



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : <b>C12N 15/00, C12P 21/06</b>	<b>A1</b>	(11) International Publication Number: <b>WO 93/04165</b> (43) International Publication Date: <b>4 March 1993 (04.03.93)</b>
(21) International Application Number: <b>PCT/US92/06549</b> (22) International Filing Date: <b>6 August 1992 (06.08.92)</b> (30) Priority data: <b>744,765</b> <b>13 August 1991 (13.08.91)</b> <b>US</b> (71) Applicant: <b>WISCONSIN MILK MARKETING BOARD</b> <b>[US/US]; 8418 Excelsior Drive, Madison, WI 53717</b> <b>(US).</b> (72) Inventors: <b>BLECK, Gregory, T. ; 411 Paunack Place, Mad-</b> <b>ison, WI 53705 (US). BREMEL, Robert, D. ; 5640 Tall</b> <b>Oaks, Waunakee, WI 53597 (US).</b> (74) Agent: <b>SARA, Charles, S.; Ross &amp; Stevens, One South</b> <b>Pinckney Street, P.O. Box 2599, Madison, WI</b> <b>53701-2599 (US).</b>	(81) Designated States: <b>AT, AU, BB, BG, BR, CA, CH, CS,</b> <b>DE, DK, ES, FI, GB, HU, JP, KP, KR, LK, LU, MG,</b> <b>MN, MW, NL, NO, PL, RO, RU, SD, SE, European pa-</b> <b>tent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT,</b> <b>LU, MC, NL, SE), OAPI patent (BF, BJ, CF, CG, CI,</b> <b>CM, GA, GN, ML, MR, SN, TD, TG).</b>  <b>Published</b> <i>With international search report.</i> <i>With amended claims.</i>	

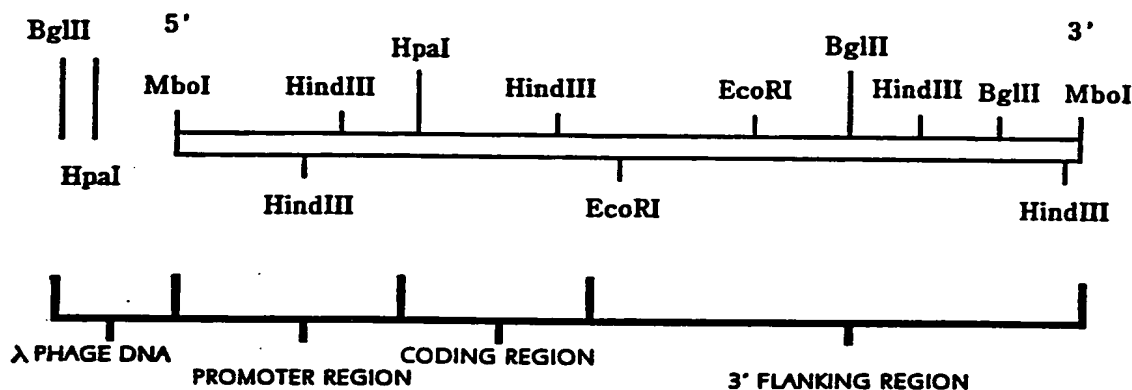
(54) Title: DNA SEQUENCE ENCODING BOVINE  $\alpha$ -LACTALBUMIN AND METHODS OF USE

## ALPHA-LACTALBUMIN GENOMIC CLONE:

2.0 kb 5' FLANKING REGION (PROMOTER REGION)

1.7 kb CODING REGION (4 EXONS)

8.8 kb 3' FLANKING REGION



## (57) Abstract

The present invention relates to the use of a mammary specific  $\alpha$ -lactalbumin protein to assist in the production of recombinant proteins in mammals' milk. The invention also relates to the genetically engineered mammal that produces the desired recombinant product in its milk and to the products produced by the genetically engineered mammal, including the altered composition of milk and the semen which includes the desired  $\alpha$ -lactalbumin protein DNA sequence.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	MN	Mongolia
AU	Australia	FR	France	MR	Mauritania
BB	Barbados	GA	Gabon	MW	Malawi
BE	Belgium	GB	United Kingdom	NL	Netherlands
BF	Burkina Faso	GN	Guinea	NO	Norway
BG	Bulgaria	GR	Greece	NZ	New Zealand
BJ	Benin	HU	Hungary	PL	Poland
BR	Brasil	IE	Ireland	PT	Portugal
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	RU	Russian Federation
CG	Congo	KP	Democratic People's Republic of Korea	SD	Sudan
CH	Switzerland	KR	Republic of Korea	SE	Sweden
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovak Republic
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CS	Czechoslovakia	LU	Luxembourg	SU	Soviet Union
CZ	Czech Republic	MC	Monaco	TD	Chad
DE	Germany	MG	Madagascar	TG	Togo
DK	Denmark	ML	Mali	UA	Ukraine
ES	Spain			US	United States of America

DNA SEQUENCE ENCODING BOVINE  $\alpha$ -LACTALBUMIN  
AND METHODS OF USE

FIELD OF THE INVENTION

The present invention relates generally to a DNA sequence encoding bovine  $\alpha$ -lactalbumin and to methods of producing proteins including recombinant proteins in the milk of lactating genetically engineered or transgenic mammals. The present invention relates also to genetically engineered or transgenic mammals that secrete the recombinant protein. The present invention is also directed to a genetic marker for identifying animals with superior milk producing characteristics.

REFERENCE TO CITED ART

Reference is made to the section preceding the CLAIMS for a full bibliography citation of the art cited herein.

DESCRIPTION OF THE PRIOR ART

$\alpha$ -Lactalbumin is a major whey protein found in cow's milk (Eigel et al., 1984). The term "whey protein" includes a group of milk proteins that remain soluble in "milk serum" or whey after the precipitation of casein, another milk protein, at pH 4.6 and 20°C.  $\alpha$ -Lactalbumin has these characteristics.

$\alpha$ -Lactalbumin is a secretory protein that normally comprises about 2.5% of the total protein in milk.  $\alpha$ -Lactalbumin has been used as an index of mammary gland function in response to hormonal regulation in bovine explant culture (Akers et al., 1981; Goodman et al., 1983) and as an index of udder development (McFadden et al., 1986).  $\alpha$ -Lactalbumin interacts with galactosyl transferase and therefore plays an essential role in the biosynthesis of milk sugar lactose (Brew, K. and R.L. Hill, 1975). Lactose is an important component in milk, and contributes to milk osmolality. It is the most constant constituent in cow's milk (Larson, 1985).  $\alpha$ -Lactalbumin is useful as an index of lactogenesis in cultured mammary tissue (McFadden et al., 1987). It is therefore believed that  $\alpha$ -Lactalbumin is an important

-2-

protein in controlling milk yield and can be used as an indicator of mammary function.

5 The expression of bovine  $\alpha$ -lactalbumin may be a potential rate limiting process in dairy cattle. If greater expression of the  $\alpha$ -lactalbumin gene can be obtained, then more milk and milk protein could be produced. In other words,  $\alpha$ -lactalbumin is a potential Quantitative Trait Locus (QTL).

#### SUMMARY OF THE INVENTION

10 One object of the present invention is to detect possible genetic differences in the expression of bovine  $\alpha$ -lactalbumin.

Another object of the present invention is to provide a DNA sequence encoding a mammary specific bovine  
15  $\alpha$ -lactalbumin protein having a specified nucleotide sequence.

It is also an object of the present invention to provide a method for genetically engineering the incorporation of one or more copies of a construct  
20 comprising an  $\alpha$ -lactalbumin control region, which construct is specifically activated in the mammary tissue.

These objects and others are addressed by the present invention, which is directed to a DNA sequence  
25 encoding bovine  $\alpha$ -lactalbumin having a specified nucleotide sequence.

The present invention is also directed to an expression vector comprising this DNA sequence. Further, the present invention is directed to the protein  $\alpha$ -  
30 lactalbumin having the nucleotide sequence.

The present invention is also directed to an expression system comprising a mammary specific  $\alpha$ -lactalbumin control region which, when genetically incorporated into a mammal, permits the female species of  
35 that mammal to produce the desired recombinant protein in its milk.

-3-

The present invention is also directed to a genetically engineered or transgenic mammal comprising the specified DNA sequence encoding bovine  $\alpha$ -lactalbumin.

5 The present invention is also directed to a DNA sequence coding for  $\alpha$ -lactalbumin, which is operatively linked to an expression system coding for a mammary-specific  $\alpha$ -lactalbumin protein control, or any control region which specifically activates  $\alpha$ -lactalbumin in milk or in mammary tissue, through a DNA sequence  
10 coding for a signal peptide that permits secretion and maturation of the  $\alpha$ -lactalbumin in the mammary tissue.

The present invention is also directed to a process for genetically engineering the incorporation of one or more copies of a construct comprising an  $\alpha$ -  
15 lactalbumin control region which specifically activates  $\alpha$ -lactalbumin in milk or in mammary tissue. The control region is operatively linked to a DNA sequence coding for a desired recombinant protein through a DNA sequence coding for a signal peptide that permits the secretion  
20 and maturation of  $\alpha$ -lactalbumin in the mammary tissue.

The present invention is also directed to a process for the production and secretion into a mammal's milk of an exogenous recombinant protein. The steps include producing milk in a genetically engineered or  
25 transgenic mammal. The milk is characterized by an expression system comprising  $\alpha$ -lactalbumin control region. The control region is operatively linked to an exogenous DNA sequence coding for the recombinant protein through a DNA sequence coding for a signal for the  
30 peptide effective in secreting and maturing the recombinant protein in mammary tissue. The milk is then collected for use. Alternatively, the exogenous recombinant protein is isolated from the milk.

35 The present invention is also directed to a selection characteristic for identifying superior milk and milk protein producing animals comprising a DNA

/\*

-4-

sequence encoding bovine  $\alpha$ -lactalbumin and having a specified nucleotide sequence.

5           The present invention is also directed to a selection characteristic for identifying superior milk and milk protein producing mammals. The mammals are characterized by inherited genetic material in the DNA structure of the mammal. The genetic material encodes at least one desired dominant selectable marker for bovine  $\alpha$ -lactalbumin. One such marker is adenosine, which is  
10       located at the -13 position on the control region of the DNA sequence for  $\alpha$ -lactalbumin. The present invention is also directed to a method of predicting superior milk and milk protein production in animals comprising identifying the selection characteristic discussed above.

15           The present invention is further directed to a method for modifying the milk composition in mammals which comprises inserting a DNA sequence encoding bovine  $\alpha$ -lactalbumin having a specified nucleotide sequence.

20           The DNA sequence and the various methods of using it have potentially beneficial uses for dairy farmers, artificial insemination organizations, genetic marker companies, and embryo transfer and cloning companies, to name a few.

25           The uses for this genetic marker include the identification of superior nuclear transfer embryos and the identification of superior embryos to clone.

30           The present invention also will aid in the progeny testing of sires. The specified DNA sequence can be used as a genetic marker to identify possible elite sires in terms of milk production and milk protein production. This will increase the reliability of buying superior dairy cattle.

35           The present invention also will provide assistance in farm management decisions, such as sire selection and selective culling. The physiological markers assist in determining future production performance in addition to a cow's pedigree. From this

-5-

information, one could buy or retain a heifer with a DNA sequence encoding  $\alpha$ -lactalbumin of the present invention and consider culling a heifer without the proper sequence.

5                    BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic illustration of a partial restriction map of the bovine  $\alpha$ -lactalbumin of the present invention. The sequence contains 2.0 kilobases of a 5' flanking region, 1.7 kilobases of a coding region and 8.8 kilobases of a 3' flanking region. Digestion with the Hpa I yields a 2.8 kilobase fragment containing the whole 5' flanking region.

10                   Fig. 2 depicts in schematic outline a map of the plasmid A-lac Pro/pIC 20R. A Hpa I fragment of the genomic clone was inserted into the EcoRV site of pIC 20R. The Hpa I fragment contains 2.1 kb of 5' flanking DNA the signal peptide coding region of  $\alpha$ -lactalbumin and 8 bases encoding the mature  $\alpha$ -lactalbumin protein. Six unique enzyme sites are available for attaching various genes to the sequence.

15                   Fig. 3 is a schematic illustration of a detailed map of the  $\alpha$ -lactalbumin 5' flanking control region cloned in EcoRV site of the plasmid pIC 20R (SEQ ID NO:1, SEQ ID NO:2).

20                   Fig. 4 is a schematic illustration of a detailed map of the 8.0 kilobase BglIII fragment.

25                   Fig. 5 depicts the nucleotide sequence (SEQ ID NO:3) of the control/enhancer region of the bovine  $\alpha$ -lactalbumin protein.

30                   Fig. 6 depicts in schematic outline a map of a plasmid containing bovine  $\alpha$ -lactalbumin-bovine  $\beta$ -casein gene construct.

35                   Fig. 7 illustrates a sequence comparison between humans and bovine genes in the 5' flanking region of the bovine  $\alpha$ -lactalbumin protein between the present invention U. S. bovine sequence (SEQ ID NO:4), a human sequence (SEQ ID NO:5) and the French bovine (SEQ ID

-6-

NO:6) for the putative steroid response element and between the present invention U. S. bovine sequence (SEQ ID NO:7), a human sequence (SEQ ID NO:8) and the French bovine (SEQ ID NO:9) for the RNA polymerase binding region, surrounding three of the four nucleotide sequence variant mutations.

Fig. 8 is a DOTPLOT™ graph comparing the bovine  $\alpha$ -lactalbumin 5' flanking sequence to the same region of the human  $\alpha$ -lactalbumin sequence.

Fig. 9 is a DOTPLOT™ graph comparing the bovine  $\alpha$ -lactalbumin 5' flanking sequence to the same region of the guinea pig  $\alpha$ -lactalbumin sequence.

Fig. 10 is a DOTPLOT™ graph comparing the bovine  $\alpha$ -lactalbumin 5' flanking sequence to the same region of the rat  $\alpha$ -lactalbumin sequence.

Fig. 11 is a graph illustrating expression levels observed in each of three  $\alpha$ -lactalbumin transgenic mouse line.

Fig. 12 is a 4% NuSieve autoradiographic gel of MnII digested PCR products.

Fig. 13 is a graph illustrating a scatter plot of each data point in Fig. 12 as well as mean values for each of the three genotypes.

#### DETAIL DESCRIPTION OF THE PREFERRED INVENTION

In the Description the following terms are employed:

Genetic engineering, manipulation or modification: the formation of new combinations of materials by the insertion of nucleic acid molecules produced outside the cell into any virus, bacterial plasmid or other vector system so as to allow their incorporation into a host organism in which they do not naturally occur, but in which they are capable of continued propagation at least throughout the life of the host organism. Although the term incorporates transgenic alteration, the manipulation of the genomic sequence does not have to be permanent, i. e., the genetic engineering



-7-

can affect only the animal which was directly manipulated.

Transgenic animals: permanently genetically engineered animals created by introducing new DNA sequences into the germ line via addition to the egg.

It is within the scope of the present application to use any mammal for the invention. Examples of mammals include cows, sheep, goats, mice, oxen, camels, water buffaloes, llamas and pigs.

Preferred mammals include those that produce large volumes of milk and have long lactating periods.

The present invention is directed to a gene which encodes bovine  $\alpha$ -lactalbumin. This gene has been isolated and characterized. The 5' flanking region of the gene has been cloned into six vectors for use as a mammary specific control region in the production of genetically engineered mammals. To better understand the regulation of this control region, 2.0 kilobases of the 5' flanking sequence have been sequenced. The  $\alpha$ -lactalbumin 5' flanking sequence serves as a useful mammary-specific "control/enhancer complex" for engineering genetic constructs that could be capable of driving the expression of novel and useful proteins in the milk of genetically engineered or transgenic mammals. This results in an increase in milk production and the protein composition in milk, a change in the milk and/or protein composition in milk, and the production of valuable proteins in the milk of genetically engineered or transgenic mammals. Such proteins include insulin, growth hormone, growth hormone releasing factor, somatostatin, tissue plasminogen activator, tumor necrosis factor, lipocortin, coagulation factors VIII and IX, the interferons, colony stimulating factor, the interleukins, urokinase, industrial enzymes such as cellulases, hemicellulases, peroxidases, and thermal stable enzymes.

-8-

The  $\alpha$ -lactalbumin gene is the preferred gene for use in the process because it is a mammary specific protein 5' control region. It also exerts the tightest lactational control of all milk proteins. Further, it is independently regulated from other milk proteins and is produced in large quantity by lactating animals.

#### Total Sequence

A gene encoding the milk protein bovine  $\alpha$ -lactalbumin was isolated from a bovine genomic library (Woychik, 1982). The Charon 28 lambda library was probed using a bovine  $\alpha$ -lactalbumin cDNA (Hurley, 1987) and a 770 base pair  $\alpha$ -lactalbumin polymerase chain reaction product. The positive lambda clone includes 12.5 kilobases of inserted bovine sequence, consisting of 2.0 kilobases of a 5' flanking (control/enhancer) region, a 1.7 kilobase coding region and 8.8 kilobases of a 3' flanking region. A partial restriction map of the clone is illustrated in Fig. 1.

A 2.8 kilobase Hpa I fragment including the 2.0 kilobase control region along with the signal peptide coding region was cloned into the EcoRV site of the plasmid pIC 20R. The plasmid is illustrated in schematic outline in Fig. 2.

An 8.0 kilobase Bgl II fragment containing a 2.0 kilobase 5' flanking control region, a 1.7 kilobase coding region, 3.0 kilobases of a 3' flanking region, 1.2 kilobases of a lambda DNA has also been isolated. Reference is made to figure 4 for a map of the 8.0 kilobase fragment. Transgenic mice have been produced using the Bgl II fragment.

#### Control/Enhancer Region

The 2.0 kilobase 5' flanking region has been cloned into the vectors Pic 20R and Bluescript KS+. A schematic illustration of the  $\alpha$ -lactalbumin 5' flanking control region cloned in the EcoRV site of pIC 20R is depicted in Figs. 2 and 3 (SEQ ID NO:1, SEQ ID NO:2).

-9-

The construct's multiple cloning site, which exists downstream of the signal peptide coding region, permits various genes to be attached to the  $\alpha$ -lactalbumin control region. Thus, this vector allows for easy attachment of specific coding sequences of genes. It contains all elements necessary for expression of proteins in milk, i.e., a mammary specific control region, a mammary specific signal peptide coding region and a mature protein-signal peptide splice site which is able to be cleaved in the mammary gland. The vector also contains many unique restriction enzyme sites for ease of cloning. Attachment of genes to this control region will allow for mammary expression of the genes when these constructs are placed into mammals. These vectors also contain the  $\alpha$ -lactalbumin signal peptide coding sequence which will allow for proper transport of the expressed protein into the milk of the lactating mammal.

The control region construct has driven mammary expression of a desired protein in transgenic mice. Bovine  $\alpha$ -lactalbumin levels of greater than 1 mg/ml have been observed in the milk of transgenic mouse lines as described in Example 2 (*infra*). Constructs containing the 2.0 kilobase region attached to the bovine  $\beta$ -casein gene (Bonsing, J., et al., 1988) as well as the bacterial reporter gene chloramphenicol acetyl transferase have been produced in our lab. Fig. 6 is a schematic representation of a plasmid containing the bovine  $\alpha$ -lactalbumin bovine  $\beta$ -casein gene construct. The genomic DNA sequence containing the bovine  $\beta$ -casein gene was attached to the 5' flanking sequence of the bovine  $\alpha$ -lactalbumin 5' flanking sequence. The vector contains the polyadenylation site of  $\beta$ -casein along with approximately 100 base pairs of 5' flanking DNA. The 100 base pairs of 5' flanking DNA is attached to the bovine  $\alpha$ -lactalbumin 5' flanking region at the -100 position. The construct uses the proximal promoter elements of  $\beta$ -casein and the distal control region elements of  $\alpha$ -

-10-

lactalbumin. The  $\beta$ -casein construct has been used to produce transgenic mice as is illustrated in the examples.

To understand the control of the control/enhancer region the 2.0 kilobases of 5' flanking region were sequenced. A single strand copy of the sequence is listed in Fig. 5 (SEQ ID NO:3). The sequence is listed 5' to 3' with the signal peptide coding region underlined.

#### 10 Regulatory Sequences

Potential regulatory sequences contained within the 5'-flanking region of bovine  $\alpha$ -lactalbumin have been identified. There are possible regulatory regions in the introns as well as in the 3' flanking region. Portions of the suspected control regions were examined for possible sequence differences in the population which might be related to milk and milk protein production of individual cows. The differences in the regulatory regions of  $\alpha$ -lactalbumin are expected to lead to differences in expression of  $\alpha$ -lactalbumin mRNA. The increased cellular content of mRNA will increase the expression of  $\alpha$ -lactalbumin protein with a concomitant increase in lactose synthase resulting, ultimately, in a milk and milk protein production increase. This type of mechanism would be considered a major gene effect on milk and milk protein production by  $\alpha$ -lactalbumin. The changes are viewed as causally-linked to changes in milk and milk protein production and not correlatively-linked. Correlatively-linked traits are those which are closely associated with an unknown genetic loci which has the direct impact on the quantitative trait.

Sequence differences between the U. S. Holstein and the French cow (Vilotte, et al., 1987) of an unknown breed were found at four positions within the 5' flanking region. One of the identified sequences has a sequence which would indicate that it was a steroid

-11-

hormone response element. Two other differences were noted in the RNA polymerase binding region and a fourth in the signal peptide coding region of the gene. Because of the relationship between these sequences and known control sequences of mammalian genes, all the variations occur in regions one would expect to be involved in regulation of the amount of mRNA produced. Further, genetic variations which occur in factors binding to these regions would also be expected to cause changes.

Fig. 7 illustrates sequence variants observed in the 5' flanking region between the present invention U. S. bovine, human (Hall et al., 1987) and the French bovine (Vilotte, 1987) for the putative steroid response element (SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6 respectively) and for the RNA polymerase binding region (SEQ ID NO:7, SEQ ID NO:8, and SEQ ID NO:9 respectively). All of the differences occur in highly conserved portions of the gene as seen by comparing this region to the same region of the human  $\alpha$ -lactalbumin gene. Fig. 7 also shows that the positions where the bovine genes differ are the same positions the human gene differs from the bovine. These data indicate that the bases are part of a potentially important control region.

A method has been devised to give a clearcut differentiation between two of the variants at a position -13 bases from the start of transcription, i. e., 13 base positions from the signal peptide coding region. The two variants are termed ( $\alpha$ -Lac (-13) A) and ( $\alpha$ -Lac (-13) B). The  $\alpha$ -lac (-13) A genotype is adenine base at position -13 the  $\alpha$ -lac (-13) B genotype is either a guanine, thymine or cytosine base at -13. They can be differentiated with a simple restriction enzyme digest of an amplified polymerase chain reaction (PCR) product using a specific restriction enzyme (MnlI). Because of the specificity of the restriction enzyme MnlI, the restriction analysis is unable to distinguish between these different possibilities. The  $\alpha$ -lac (-13) A allele

-12-

contains an extra MnlI site at position -13 giving the smaller band observed on the gel.

To amplify the appropriate region of DNA, oligonucleotides which frame the sequence of interest were synthesized. These oligonucleotides were chosen because of their specific chemical characteristics. These oligonucleotides were then used in a polymerase chain reaction to amplify the framed portion of the  $\alpha$ -lactalbumin gene. The oligonucleotides have the following sequences:

$\alpha$ -lac Seq. 1 (SEQ ID NO:10)

5'ACGCTTGTAACGACGGCCAGTTGATTCTCAGTCTCTGTGGT 3'

$\alpha$ -lac Seq. 2 (SEQ ID NO:11)

5'AGCATCAGGAAACAGCTATGACCTGGGTGGCATGGAATAGGAT 3'

Restriction fragment analysis (Sambrook, J. et al., 1989) was used to examine animals from a number of breeds of cattle. In most breeds, namely, Jersey, Guernsey, Brown Swiss, Simmental and Brahman, only one of two genotypes is found. This is the  $\alpha$ -lac (-13) B genotype. However, in the most popular and highest milk producing breed of cattle, the Holstein, two genotypes occur at this position. The frequency of the A genotype was 27% in random samples, while the frequency of the B genotype was 73%. Holsteins contain both the genotype found in the other breeds as well as a separate distinct genotype which appears to have arisen within the last thirty years in the U. S. Holstein population as determined by examining pedigrees of sires currently in use. It appears that this genotype has unknowingly been selected for using traditional animal selection. Homozygous and heterozygous animals are found within the Holstein population.

The genotype ( $\alpha$ -lac (-13)) has been examined for its correlation with milk and milk protein production. The three additional variations are being examined to determine the frequency of their differences in the cattle population and their correlation with milk

-13-

and milk protein production. The possible linkage of these genotypes is also being examined using DNA sequencing. The goal of this technology is to identify the optimal regulatory genotype for  $\alpha$ -lactalbumin and to select animals with those particular characteristics.

#### Detection and Selection of Four Genetic Variants

The region of sequence where the  $\alpha$ -lac (-13) variation occurs can be amplified using the polymerase chain reaction (PCR) (Sambrook et al., 1989) and two of the following primers which were developed. Each primer allows for amplification of a specific portion of the  $\alpha$ -lactalbumin gene. Combinations of the listed primers can be used in between any two of the primer locations listed below.

Primer No. \ (SEQ ID NO:)	Primer sequence	Primer location (From translation start site)
1 \ (12)	5' CTCTTCCTGGATGTAAGGCTT 3'	(-120)-(-100)
2 \ (13)	5' TCCTGGGTGCTCATTGAAAGGACT 3'	(-2000)-(-1975)
3 \ (14)	5' CAATGTGGTATCTGGCTATTTAGTG 3'	(-717)-(-692)
4 \ (15)	5' AGCCTGGGTGGCATGGAATA 3'	(+53)-(+33)
5 \ (16)	5' GAAACGCGGTACAGACCCCT 3'	(+453)-(+433)

After amplification of the specific region, the DNA is either sequenced or digested with restriction enzymes to detect the sequence differences. In the case of the  $\alpha$ -lac (-13) variation, the sequence difference can be seen using the restriction enzyme MnlI (5'CTCC 3' recognition site). The PCR DNA product is digested with MnlI and then run on a 4% NuSieve agarose gel to observe the polymorphism.

A 650 base pair sequence containing all four of the variations is being examined using a unique sequencing technique. PCR is initially used to amplify a 770 base pair portion of the  $\alpha$ -lactalbumin 5' flanking region. Another PCR reaction is then performed using a portion of the initial reaction and the following primers (SEQ ID NO:10 and SEQ ID NO:11 respectively):

$\alpha$ -lac seq. 1 5'ACGCTTGTAACGACGGCCAGTTGATTCTCAGTCTCTGTGGT 3'

-14-

$\alpha$ -lac seq. 2 5'AGCATCAGGAAACAGCTATGACCTGGGTGGCATGGAATAGGAT 3'

The primers listed above contain a portion of the  $\alpha$ -lactalbumin gene as well as both M13 DNA sequencing primers. The primers are designed to allow for DNA sequencing to be performed in both directions on the PCR DNA product. The final PCR product will contain the region of  $\alpha$ -lactalbumin containing the four genetic variants, the two M13 sequencing priming regions and 5 "dummy bases" on the end to aid in the M13 primer binding.

#### Comparison of Highly Conserved Portions of the 5' Flanking Region of $\alpha$ -Lactalbumin Between Species

Reference is made to Figs. 8 - 10 for DOTBLOT™ graphs comparing the bovine  $\alpha$ -lactalbumin sequence to the same region of the human (Fig. 8), guinea pig (Fig. 9), and rat (Fig. 10). The region in Fig. 8 (human) spans 819 base pairs. The sequences are highly conserved to about 700 base pairs. The region in Fig. 9 (guinea pig) spans 1381 base pairs. The sequences are highly conserved to about 700 base pairs, but then diverge. The region in Fig. 10 (rat) spans 1337 base pairs. The sequences are highly conserved to about 700 base pairs, but then diverge. Species differences in control regions would be expected to occur in non-conserved regions of the sequence.

#### Comparison of 5' Flanking Region of Bovine $\alpha$ -Lactalbumin to Other Bovine Milk Protein Genes

Portions of the 5' flanking region of the other bovine milk protein genes ( $\alpha$ s1 and  $\alpha$ s2 casein,  $\beta$ -casein, K-casein and  $\beta$ -lactoglobulin) which are highly conserved with the  $\alpha$ -lactalbumin 5' flanking region were identified. It is probable that sequence differences within these regions will also have an effect on mRNA production as well as final protein production. Two examples of these highly homologous regions are listed below.

The bovine  $\alpha$ -lactalbumin sequence from (-161) - (-115) (SEQ ID NO: 17) compared to the bovine  $\beta$ -casein



-15-

sequence (SEQ ID NO:18) corresponding to the same region of the gene. Percent similarity is 69% over 46 bases.

```

5      AGGAAGCTCAATGTTTCTTTGTTGGTTTTACTGGCCTCTCTTGTC
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      AGGAGGCT.ATTCTTTCCTTTTAGTCTATACTGTCTTCGCTCTTCA

```

10 The bovine  $\alpha$ -lactalbumin sequence (SEQ ID NO:19) from (-1420) - (-1351) is compared to the bovine  $\beta$ -casein sequence (SEQ ID NO:20) corresponding to the same region of the gene. Percent similarity is 75% over 69 bases.

```

      TATAAGAAATCAGGCTTTAGAGACTGATGTAGAGAGAATGAGCCCTGGCA
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      TCTCAGAAATCACACTTTTTTGCCTGTG.....GCCTTGGA

```

```

15      TACCAGAAGCTAACAGCTA
      ||| ||||| ||| |||
      .ACCAAAGCTAACACATA

```

20 The included data indicate that the bovine  $\alpha$ -lactalbumin gene will be useful as selection tool in the dairy cattle industry as well as a valuable control/enhancer and gene to be used in the field of genetically engineered mammals. The control region we have cloned contains the necessary regulatory elements to express genes in the milk of genetically engineered mammals as well as the "high expressing genotype" as shown by our milk and milk protein production and sequence variation data. These facts make this a useful gene in both industrial and research areas. Application of these techniques to the other milk proteins will allow for the selection of valuable genotypes corresponding to the  $\beta$ -casein,  $\alpha_1$ - and  $\alpha_2$ -casein and K-casein genes and the  $\beta$ -lactoglobulin genes.

#### Coding Region

35 The coding region of the  $\alpha$ -lactalbumin protein includes a 1.7 kilobase sequence.

#### 3' Flanking Region

The 3' flanking region is an 8.8 kilobase flanking region downstream of the DNA sequence coding for the desired recombinant protein. This region apparently

-16-

stabilizes the RNA transcript of the expression system and thus increases the yield of desired protein from the expression system.

#### Operation

5                   The above-described expression systems may be prepared by methods well-known in the art. Examples include various ligation techniques employing conventional linkers, restriction sites, etc. Preferably, these expression systems are part of larger  
10                   plasmids.

                  After isolation and purification, the expression systems or constructs are added to the gene pool which is to be genetically altered.

                  The methods for genetically engineering  
15                   mammals are well-known to the art. Reference is made to to Alberts, B. et al., 1989 and Lewin, B. 1990, for textbook descriptions of genetic engineering and transgenic alteration of animals. Briefly, genetic engineering involves the construction of expression  
20                   vectors so that a cDNA clone or genomic structure is connected directly to a DNA sequence that acts as a strong promoter for DNA transcription. By means of genetic engineering, mammalian cells, such as mammary tissue, can be induced to make vast quantities of useful  
25                   proteins.

                  For the purposes of this invention, the term "genetic engineering," as defined supra. in the list of definitions, includes single line alteration, i. e., genetic alteration only during the life of the affected  
30                   animal with no germ line permanence. The construct can be genetically incorporated in mammalian glands such as mammary glands and mammalian stem cells.

                  Genetic engineering also includes transgenic alteration, i. e. the permanent insertion of the gene  
35                   sequence into the genomic structure of the affected animal and any offspring. Transgenically altering a mammal involves microinjecting a DNA construct into the

-17-

pronuclei of the fertilized mammalian egg to cause one or more copies of the construct to be retained in the cells of the developing mammal. In a transgenic animal, the engineered genes are permanently inserted into the germ line of the animal.

The genetically engineered mammal is then characterized by an expression system comprising the  $\alpha$ -lactalbumin control region operatively linked to an exogenous DNA sequence coding for the recombinant protein through a DNA sequence coding for a signal peptide effective in secreting and maturing the recombinant protein in mammary tissue. In order to produce and secrete the recombinant protein into the mammal's milk, the transgenic mammal must be allowed to produce the milk, after which the milk is collected. The milk may then be used in standard manufacturing processes. The exogenous recombinant protein may also be isolated from the milk according to methods known to the art.

#### Selection Characteristics

The  $\alpha$ -lactalbumin control/enhancer sequence of Fig. 1 is also important as a selection characteristic for identifying superior or elite milk producing mammals. Presently, those in the dairy cattle business can only rely on pedigree information, which is frequently not available, to predict milk and milk protein production in mammals, specifically the bovine species. The study of physiological markers as a means for determining milk and milk protein production has received some interest. The most common physiological marker traits studied in dairy cattle are hormones, enzymes, and different blood metabolites. Components of the immune system have also been studied. Traits listed as possible marker traits for milk yield include thyroxine, blood urea nitrogen, growth hormones, insulin-like growth factors and insulin, and glucose and free fatty acids. While these techniques have shown some advances in predicting milk and milk

-18-

protein production in a dairy animal, there is currently no other reliable means to predict these characteristics.

The present invention provides a selection characteristic for identifying superior milk and milk protein-producing mammals comprising inherited genetic material which is DNA occurring in the genetic structure of the mammal in which the genetic material encodes a dominant selectable marker for bovine  $\alpha$ -lactalbumin.

The DNA sequence disclosed herein serves as a characteristic marker for elite milk producing mammals.

The examples below describe the invention disclosed herein, although the invention is not to be understood as limited in any way to the terms and scope of the examples.

#### EXAMPLES

Example 1:  $\alpha$ -lac (-13) variation study.

Forty-two mammals were selected in a stratified random manner to provide mammals of a wide range of milk and milk protein production capabilities within the UW herd.

DNA was isolated according to procedures known to the art from a random sample of 42 Holstein dairy cows in the University of Wisconsin-Madison herd. Each mammal was genotyped as described previously for the  $\alpha$ -lactalbumin (-13) variation using a 4% NuSieve gel of MnlI digested PCR products.

The gene frequency in this population is 28% for the  $\alpha$ -lac (-13) A and 72% for the  $\alpha$ -lac (-13) B. Each of the distinct genotypes are shown on the gel in Fig. 12. The legend for the gel of Figure 12 is as follows:

Lane 1	Molecular Weight Standards	
Lane 2-3	heterozygous	$\alpha$ -lac (-13) AB
Lane 4:	homozygous	$\alpha$ -lac (-13) BB
Lane 5	heterozygous	$\alpha$ -lac (-13) AB
Lane 6	homozygous	$\alpha$ -lac (-13) BB
Lane 7	homozygous	$\alpha$ -lac (-13) AA

-19-

Lane 8 heterozygous  $\alpha$ -lac (-13) AB

Analysis of the genetic capabilities of the 42 mammals indicates a possible major gene effect caused by the  $\alpha$ -lac (-13) allele or linked to the  $\alpha$ -lac (-13) allele. A scatter plot of each data point as well as mean values for each of the three genotypes is illustrated in Fig. 13. Holstein cows were compared using their predicted transmitting ability for milk.

The data indicate that the  $\alpha$ -lac (-13)A genotype is the preferred genotype for milk and milk protein production. Table 1 shown below indicates the statistical association of differences in milk and milk protein production ability observed between each of the genotypes for the traits listed below. Analysis of variance and T tests (LSD) were performed on the data. All of the production yield traits were positively correlated with the  $\alpha$ -lac (-13) A allele. Milk protein percentage was negatively correlated to the  $\alpha$ -lac (-13) A allele.

Table 1

Trait/Genotype	Genotype		
	$\alpha$ -lac (-13) AA	$\alpha$ -lac (-13) AB	$\alpha$ -lac (-13) BB
PTA (Milk) /AA	-----	N.S.	p<0.02
PTA (Milk)/AB	N.S.	-----	p<0.02
ME305 (Milk)/AA	-----	N.S.	N.S.
ME305 (Milk)/AB	N.S.	-----	p<0.1
PTA (Protein #)/AA	-----	N.S.	N.S.
PTA (Protein #)/AB	N.S.	-----	p<0.1
PTA (Protein %)/AA	-----	N.S.	p<0.01
PTA (Protein %)/AB	N.S.	-----	p<0.01

Example 2. Production of Transgenic mice to study the regulation of bovine  $\alpha$ -lactalbumin gene expression.

Genomic Library Screening:

The gene encoding the milk protein bovine  $\alpha$ -lactalbumin was isolated from a bovine genomic library (Woychik, 1982). The genomic library was screened

-20-

according to the following procedure. Approximately 1.5 million lambda plaques were transferred to nylon membranes using procedures described by Maniatis et al. (1989). The  $\alpha$ -lactalbumin cDNA (Hurley, 1987) or a 770 base pair PCR product was nick translated (BRL) with a-P32 labeled dCTP. Blots were prehybridized overnight (65C) then hybridized for 16 hours at 65C. Blots were washed (Twice in 2X SSC 1% SDS, Once in 0.1X SSC 0.1% SDS) at 65C and placed on Kodak X-OMAT film for autoradiography. A 8.0 kilobase fragment containing the  $\alpha$ -lactalbumin gene was purified as illustrated in Fig. 4. The 8.0 kilobase fragment contained 2.1 kilobases of 5' flanking region, the 1.7 kilobase coding region and 2.6 kilobases of 3' flanking region.

#### Production of transgenic mice:

Mature C57B6 X DBA2J F1 (B6D2) female were superovulated (PMSG and hCG) and mated with ICR or B6D2 males to yield fertilized eggs for pronuclear microinjection. The eggs were microinjected using a Leitz micromanipulator and a Nikon inverted microscope. Forty normal appearing two cell embryos were transferred to each pseudopregnant recipient. (University of Wisconsin-Madison Biotechnology Center Transgenic Mouse Facility, Dr. Jan Heideman).

#### Screening of transgenic mice using PCR:

Tail DNA was extracted using the method described by Constantini et al. (1986). Polymerase chain reaction (PCR) was performed using 10 ml 10x PCR reaction buffer (Promega Corp., Madison, WI.), 200 mM each dNTP (Pharmacia Intl., Milwaukee, WI.), 1.0  $\mu$ M each primer (upstream primer 25mer -712 to -687 (5' CAATGTGGTATCTGGCTATTTAGTG 3') (SEQ ID NO:14), downstream primer 20mer +39 to +59 (5' AGCCTGGGTGGCATGGAATA 3') (SEQ ID NO:15), 1 unit Taq DNA polymerase (Promega Corp., Madison, WI.) and 1mg genomic DNA. Volume was adjusted to 100 ml with double distilled sterile water and reaction was overlaid with heavy mineral oil. Samples

-21-

were subjected to 30 cycles (94C 2 min., 50C 1.5 min., 72C 1.5 min.). Products were run in an 1% agarose gel and stained with ethidium bromide.

Mouse Milking:

5                   The mice were separated from their litters for four hours and then anesthetized (0.01 ml/g body weight I.P. injection of 36% propylene glycol, 10.5% ethyl alcohol (95%), 41.5% sterile water, and 12% sodium pentobarbitol (50 mg/ml)). After being anesthetized the  
10                   mice were injected I.M. with 0.3 I.U. oxytocin and milked using a small vacuum milking machine. Three of fifty-one live offspring were identified as being transgenic using polymerase chain reaction. Reference is made to Fig. 14 for a graph illustrating expression levels observed in  
15                   each of the 3  $\alpha$ -lactalbumin transgenic mouse line.

ELISA:

                  Second generation mammals from one line were milked and analysis was performed using an ELISA (enzyme linked immunosorbent assay) for bovine  $\alpha$ -lactalbumin  
20                   according to the following procedure:

1. Coat 1/40k bovine  $\alpha$ -lactalbumin antiserum 100 ml per well (in 0.05M carbonate buffer, pH 9.6) on Nunc-Immuno Plate IF MaxiSorp.
2. Wash 4x with wash buffer (0.025% Tween 20 in PBS pH 7.2)  
25
3. Add 50 ml assay buffer (0.04M MOPS, 0.12M NaCl, 0.01M EDTA, 0.1% gelatin, 0.05% Tween 20, 0.005% chlorhexidine digluconate, Leupeptin 50 mg/ml, pH 7.4).
- 30                   4. Add 50 ml of standards and samples (in assay buffer) in triplicate.
5. Add 50 ml 1/100k diluted  $\alpha$ -lactalbumin biotin conjugate.
6. Incubate overnight at 4C
- 35                   7. Wash 4x with wash buffer
8. Add 100 ml 1/10k assay buffer diluted ExtrAvidin-peroxidase (Sigma). Incubate 2 hours at RT.

-22-

9. Wash 4x twice with wash buffer.

10. Add 125 ml fresh substrate buffer (200 ml tetramethylbenzidine 20 mg/ml) DMSO, 64 ml 0.5M hydrogen peroxide, 19.74 ml sodium acetate, pH 4.8).

5 11. Incubate for 12 minutes at RT.

12. Add 50 ml 0.5M sulfuric acid to stop substrate reaction.

13. Read absorbance at 450 nm minus 600 nm in an EIA autoreader.

10 Bovine  $\alpha$ -lactalbumin was present at a concentration of levels up to and beyond 1.0 mg/ml mouse milk. Expression was determined by Western Blotting in the following steps. The 14% PAGE gel was transferred to an Immobilon-P membrane (Millipore), which was blocked in  
15 0.02 M sodiumphosphate, 0.12M NaCl, 0.01% gelatin, 0.05% Tween 20, pH=7.2, and incubated with anti-bovine  $\alpha$ -lactalbumin (1/2000 dilution) for 2 hours at room temperature. The gel was washed twice (2 min.) with an  
20 ELISA wash buffer and incubated with goat anti-rabbit IgG-HRP for 2 hours at room temperature, followed by washing 3 times with a wash buffer and washing once with double-distilled water. The gel was placed in a substrate solution (25 mg 3,3'-diaminobenzidine, 1 ml 1% CoCl<sub>2</sub> in H<sub>2</sub>O, 49 ml PBS pH 7.4 and 0.05 ml 30% H<sub>2</sub>O<sub>2</sub>) and monitored  
25 for color development. The membrane was air dried.

It is understood that the invention is not confined to the particular constructions and arrangements herein illustrated and described, but embraces such modified forms thereof as come within the scope of the  
30 claims following the bibliography.

#### BIBLIOGRAPHY OF CITED REFERENCES

Akers, R. M. et al., 1981, "Prolactin regulation of milk secretion and biochemical differentiation of mammary epithelial cells in  
35 periparturient cows." Endocrinology, 109:23.



-23-

Alberts, B. et al., 1989, Molecular Biology of The Cell (Second Edition), Garland Publishing, Inc., New York, pp. 265-271.

5        Bonsing, J. et al., 1988, "Complete nucleotide sequence of the bovine beta-casein gene," Aust. J. Biol. Sci., 41: 527-537.

Brew, K. and R. L. Hill, 1975, "Lactose biosynthesis." Rev. Physiol. Biochem. Pharmacol., 72:105.

10        Eigel, W.N. et al., 1984, "Nomenclature of proteins of cow's milk: fifth revision." J. Dairy Sci., 67:1599.

15        Goodman, G. T. et al., 1983, "Hormonal regulation of alpha-lactalbumin secretion from bovine mammary tissue cultured in vitro." Endocrinology, 112:1324.

Hall, L., et al., 1987, "Organization and sequence of the human  $\alpha$ -lactalbumin and the origins of lactation," Biochem. J., 242: 735-742.

20        Hurley, W. L. and L. A. Schuler, 1987, "Molecular cloning and nucleotide sequence of a bovine  $\alpha$ -lactalbumin cDNA," Gene, 61: 119-122.

25        Larson, B. L., 1985, "Biosynthesis and cellular secretion of milk." In: Lactation, pp. 129-163, edited by B. L. Larson, The Iowa State University Press, Ames.

Lewin, B., 1990, GENES IV, Oxford University Press, New York, pp. 691-702.

30        McFadden, T.B. et al., 1987, "Alpha-lactalbumin in bovine serum: relationships with udder development and function." J. Dairy Sci., 70:259.

Sambrook, J. et al., 1989, Molecular Cloning - A Laboratory Manual (Second Edition), Cold Spring Harbor Laboratory Press.

35        Vilotte, J. et al., 1987, "Complete nucleotide sequence of bovine  $\alpha$ -lactalbumin gene:

-24-

comparison with its rat counterpart. Biochimie, 69: 609-620.

Woychik, R., et al., Nucl. Acids Res., 10:7197-7210 (1982).

-25-

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: BLECK, GREGORY T.  
BREMEL, ROBERT D.

(ii) TITLE OF INVENTION: DNA SEQUENCE ENCODING BOVINE  
ALPHA-LACTALBUMIN AND METHODS OF USE

(iii) NUMBER OF SEQUENCES: 20

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: ANDRUS, SCEALES, STARKE & SAWALL  
(B) STREET: 100 E. WISCONSIN AVE., SUITE 1100  
(C) CITY: MILWAUKEE  
(D) STATE: WI  
(E) COUNTRY: USA  
(F) ZIP: 53202-4178

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Sara, Charles S  
(B) REGISTRATION NUMBER: 30,492  
(C) REFERENCE/DOCKET NUMBER: F. 3262-1

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (608) 255-2022  
(B) TELEFAX: (608) 255-2182  
(C) TELEX: 26832 ANDSTARK

## (2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGACCATGA TTACGAATTC ATCGTA

26

## (2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 88 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

-26-

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GAACAGTTAT CTAGATCTCG AGCTCGCGAA AGCTTGCATG CCTGCAGGTC GACTCTAGAG 60  
GATCCCCGGG TACCGAGCTC GAATTCAC 88

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2044 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: signal peptide coding region  
(B) LOCATION: 1943..2043

(ix) FEATURE:

(A) NAME/KEY: inherited control region for a-lactalbumin  
(B) LOCATION: 1966

(ix) FEATURE:

(A) NAME/KEY: putative steroid response element  
(B) LOCATION: 1433..1446

(ix) FEATURE:

(A) NAME/KEY: RNA polymerase binding region  
(B) LOCATION: 1961..1978

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GATCAGTCCT GGGTGGTCAT TGAAAGGACT GATGCTGAAG TTGAAGCTCC AATACTTTGG 60  
CCACCTGATG CGAAGAACTG ACTCATGTGA TAAGACCCTG ATACTGGGAA AGATTGAAGG 120  
CAGGAGGAGA AGGGATGACA GAGGATGGAA GAGTTGGATG GAATCACCAA CTCGATGGAC 180  
ATGAGTTTGA GCAAGCTTCC AGGAGTTGGT AATGGGCAGG GAAGCCTGGC GTGCTGCAGT 240  
CCATGGGGTT GCAAAGAGTT GGACACTACT GAGTGACTGA ACTGAACTGA TAGTGTAATC 300  
CATGGTACAG AATATAGGAT AAAAAAGAGG AAGAGTTTGC CCTGATTCTG AAGAGTTGTA 360  
GGATATAAAA GTTTAGAATA CCTTTAGTTT GGAAGTCTTA AATTATTTAC TTAGGATGGG 420  
TACCCACTGC AATATAAGAA ATCAGGCTTT AGAGACTGAT GTAGAGAGAA TGAGCCCTGG 480  
CATACCAGAA GCTAACAGCT ATTGGTTATA GCTGTTATAA CCAATATATA ACCAATATAT 540  
TGGTTATATA GCATGAAGCT TGATGCCAGC AATTGAAGG AACCATTTAG AACTAGTATC 600  
CTAAACTCTA CATGTTCCAG GACACTGATC TTAAAGCTCA GGTTCAGAAT CTTGTTTTAT 660  
AGGCTCTAGG TGTATATTGT GGGGCTTCCC TGGTGGCTCA GATGGTAAAG TGTCTGCCTG 720  
CAATGTGGGT GATCTGGGTT CGATCCCTGG CTTGGGAAGA TCCCCTGGAG AAGGAAATGG 780  
CAACCCACTC TAGTACTCTT ACCTGGAAAA TTCCATGGAC AGAGGAGCCT TGTAAGCTAC 840

-27-

	AGTCCATGGG ATTGCAAAGA GTTGAACACA ACTGAGCAAC TAAGCACAGC ACAGTACAGT	900
	ATACACCTGT GAGGTGAAGT GAAGTGAAGG TTCAATGCAG GGTCTCCTGC ATTGCAGAAA	960
	GATTCTTTAC CATCTGAGCC ACCAGGGAAG CCCAAGAATA CTGGAGTGGG TAGCCTATTC	1020
	CTTCTCCAGG GGATCTTCCC ATCCCAGGAA TTGAACTGGA GTCTCCTGCA TTTCAGGTGG	1080
5	ATTCTTCACC AGCTGAACTA CCAGGTGGAT ACTACTCCAA TATTAAAGTG CTAAAGTCC	1140
	AGTTTTCCCA CCTTTCCCAA AAAGGTTGGG TCACTCTTTT TTAACCTTCT GTGGCCTACT	1200
	CTGAGGCTGT CTACAAGCTT ATATATTTAT GAACACATTT ATTGCAAGTT GTTAGTTTTA	1260
	GATTTACAAT GTGGTATCTG GCTATTTAGT GGTATTGGTG GTTGGGGATG GGGAGGCTGA	1320
	TAGCATCTCA GAGGGCAGCT AGATACTGTC ATACACACTT TTCAAGTTCT CCATTTTTGT	1380
10	GAAATAGAAA GTCTCTGGAT CTAAGTTATA TGTGATTCTC AGTCTCTGTG GTCATATTCT	1440
	ATTCTACTCC TGACCACTCA ACAAGGAACC AAGATATCAA GGGACACTTG TTTTGTTC	1500
	TGCCTGGGTT GAGTGGGCCA TGACATATGA TGATGTACAG TCCTTTTCCA TATTCTGTAT	1560
	GTCTCTAAGA GGAAGGAGGA GTTGGCCGTG GACCCTTTGT GCATTTTCTG ATTGCTTCAC	1620
	TTGTATTACC CCTGAGGCCC CCTTTGTTCC TGAAATAGGT TGGGCACATC TTGCTTCCTA	1680
15	GAACCAACAC TACCAGAAAC AACATAAATA AAGCCAAATG GGAAACAGGA TCATGTTTGT	1740
	AACACTCTTT GGGCAGGTAA CAATACCTAG TATGGACTAG AGATTCTGGG GAGGAAAGGA	1800
	AAAGTGGGGT GAAATTACTG AAGGAAGCTC AATGTTTCTT TGTGGGTTTT ACTGGCCTCT	1860
	CTTGTCATCC TCTTCCTGGA TGTAAGGCTT GATGCCAGGG CCCCTAAGGC TTTTCCACA	1920
	AATAAAAGGA GGTGAGCAGT GTGGTGACCC CATTTCAGAA TCTTGAGGGG TAACCAAAT	1980
20	GATGTCCTTT GTCTCTCTGC TCCTGGTAGG CATCCTATTC CATGCCACCC AGGCTGAACA	2040
	GTTA	2044

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CATATTCTAT TCTA

14

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

-28-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CATATTCTAT TCCTA

15

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CATATTCTAT TTCTA

15

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TCTTGAGGGG TAACCAA

18

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TCTTGGGGGT AGCCAA

17

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

-29-

TCTTGGGGGG TCACCAA

18

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ACGCTTGTA A ACGACGGCC AGTTGATTCT CAGTCTCTGT GGT

43

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AGCATCAGGA AACAGCTATG ACCTGGGTGG CATGGAATAG GAT

43

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CTCTTCCTGG ATGTAAGGCT T

21

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TCCTGGGTGG TCATTGAAAG GACT

24

(2) INFORMATION FOR SEQ ID NO:14:

-30-

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:  
CAATGTGGTA TCTGGCTATT TAGTG
- (2) INFORMATION FOR SEQ ID NO:15:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:  
AGCCTGGGTG GCATGGAATA
- (2) INFORMATION FOR SEQ ID NO:16:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:  
GAAACGCGGT ACAGACCCCT
- (2) INFORMATION FOR SEQ ID NO:17:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 46 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:  
AGGAAGCTCA ATGTTTCTTT GTTGGTTTTA CTGGCCTCTC TTGTCA
- (2) INFORMATION FOR SEQ ID NO:18:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 45 base pairs  
(B) TYPE: nucleic acid



-31-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

5 AGGAGGCTAT TCTTTCCTTT TAGTCTATAC TGTCTTCGCT CTTC

45

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

15 TATAAGAAAT CAGGCTTTAG AGACTGATGT AGAGAGAATG AGCCCTGGCA TACCAGAAGC  
TAACAGCTA

60

69

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 55 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TCTCAGAAAT CACACTTTTT TGCCTGTGGC CTTGGCAACC AAAAGCTAAC ACATA

55

-32-

CLAIMS

What is claimed is:

1. A mammary specific DNA sequence encoding bovine  $\alpha$ -lactalbumin and promoting quantitative differences in gene expression among mammals, wherein the DNA sequence is characterized by variations in the gene structure in the control region of bovine  $\alpha$ -lactalbumin.
2. The DNA sequence of claim 1 wherein one of the variations is in the -13 position of the DNA sequence.
3. The DNA sequence of claim 2 wherein the -13 position is occupied by adenine.
4. The DNA sequence of claim 1 comprising the following DNA sequence (SEQ ID NO:19) in the control region of bovine  $\alpha$ -lactalbumin:  
TATAAGAAATCAGGCTTTAGAGACTGATGTAGAGAGAATGAGCCCTGGCATAACCAGA  
AGCTAACAGCTA.
5. The DNA sequence of claim 1 comprising the following DNA sequence (SEQ ID NO:17) in the control region of bovine  $\alpha$ -lactalbumin:  
AGGAAGCTCAATGTTTCTTTGTTGGTTTTACTGGCCTCTCTTGTC.
6. The DNA sequence of claim 1 comprising the DNA sequence listed in Fig. 5 (SEQ ID NO:3) in the control region of bovine  $\alpha$ -lactalbumin.
7. An expression vector comprising the DNA sequence of claim 1.
8. An expression system comprising a mammary specific  $\alpha$ -lactalbumin control region construct which when genetically incorporated into a mammal permits the female species of that mammal to produce the desired recombinant protein in its milk.
9. The expression system of claim 8 which comprises at least one  $\alpha$ -lactalbumin control region construct operatively linked to a DNA sequence coding for a signal peptide and  $\alpha$ -lactalbumin.

-33-

10. The expression system of claim 8 which comprises a 3' flanking region downstream of the DNA sequence coding for  $\alpha$ -lactalbumin.

11. The expression system of claim 10 wherein the construct includes a 5' flanking region upstream of the DNA sequence coding for the signal peptide.

12. The expression system of claim 8 wherein the construct comprises a 5'  $\alpha$ -lactalbumin flanking region attached to a bovine  $\beta$ -casein gene.

13. The expression system of claim 12 wherein the construct contains the polyadenylation site of  $\beta$ -casein and approximately 100 base pairs of 5'  $\alpha$ -lactalbumin flanking region.

14. The expression system of claim 12 wherein the construct includes the proximal promoter from a first milk protein and the distal control region of a second milk protein.

15. The expression system of claim 14 wherein the first and second milk proteins are selected from the group consisting of  $\alpha$ -lactalbumin,  $\beta$ -casein,  $\alpha_1$ -casein,  $\alpha_2$ -casein and K-casein.

16. The expression system of claim 12 wherein the construct includes the proximal promoter of  $\beta$ -casein and the distal control region of  $\alpha$ -lactalbumin.

17. A genetically engineered mammal characterized by an expression system comprising an  $\alpha$ -lactalbumin control region operatively linked to an exogenous DNA sequence coding for a desired protein to be expressed in milk through a DNA sequence coding for a signal peptide effective in secreting and maturing the protein in mammary tissue.

18. The genetically engineered mammal of claim 17 wherein the  $\alpha$ -lactalbumin control region includes a mammary specific DNA sequence encoding bovine  $\alpha$ -lactalbumin and having the following nucleotide sequence (SEQ ID NO:19) in the control region of bovine  $\alpha$ -lactalbumin:

-34-

TATAAGAAATCAGGCTTTAGAGACTGATGTAGAGAGAATGAGCCCTGGCATAACCAGA  
AGCTAACAGCTA.

19. The genetically engineered mammal of claim 17 wherein the  $\alpha$ -lactalbumin control region includes a mammary specific DNA sequence encoding bovine  $\alpha$ -lactalbumin and having the following nucleotide sequence (SEQ ID NO:20) in the control region of bovine  $\alpha$ -lactalbumin:

AGGAAGCTCAATGTTTCTTTGTTGGTTTTACTGGCCTCTCTTGTC.

20. Products produced by the genetically engineered mammal of claim 17.

21. Semen produced by the genetically engineered mammal of claim 17.

22. Milk produced by the genetically engineered mammal of claim 17.

23. A transgenic mammal of claim 17.

24. A DNA sequence coding for  $\alpha$ -lactalbumin which is operatively linked in an expression system of a mammary specific  $\alpha$ -lactalbumin protein control region, or any control region which specifically activates  $\alpha$ -lactalbumin in milk or in mammary tissue, through a signal peptide that permits secretion and maturation of the  $\alpha$ -lactalbumin in the mammary tissue.

25. The DNA sequence of claim 24 comprising the following DNA sequence (SEQ ID NO:19) in the control region of bovine  $\alpha$ -lactalbumin:

TATAAGAAATCAGGCTTTAGAGACTGATGTAGAGAGAATGAGCCCTGGCATAACCAGA  
AGCTAACAGCTA.

26. The DNA sequence of claim 24 comprising the following DNA sequence (SEQ ID NO:20) in the control region of bovine  $\alpha$ -lactalbumin:

AGGAAGCTCAATGTTTCTTTGTTGGTTTTACTGGCCTCTCTTGTC.

27. The DNA sequence of claim 24 comprising the DNA sequence (SEQ ID NO:3) listed in Fig. 5 in the control region of bovine  $\alpha$ -lactalbumin.

28. A process for genetically engineering the incorporation of one or more copies of a construct

-35-

comprising a  $\alpha$ -lactalbumin control region or any control region sequence specifically activated in mammary tissue, operatively linked to a DNA sequence coding for a desired recombinant protein through a DNA sequence coding for a signal peptide that permits the secretion and maturation of  $\alpha$ -lactalbumin in the mammary tissue

29. The process of claim 28 wherein the construct is genetically incorporated into a mammal and the recombinant protein product is subsequently expressed and secreted into or along with the milk of the lactating genetically engineered mammal.

30. The process of claim 29 wherein the construct is generally incorporated in mammalian embryos, mammalian mammary glands, or mammalian stem cells.

31. The process of claim 28 wherein the mammals are cows; sheep, goats, mice, oxen, camels, water buffaloes, llamas and pigs.

32. A process for the production and secretion into mammal's milk of an exogenous recombinant protein comprising the steps of:

- a. producing milk in a genetically engineered mammal characterized by an expression system comprising  $\alpha$ -lactalbumin control region operatively linked to an exogenous DNA sequence coding for the recombinant protein through a DNA sequence coding for a signal peptide effective in secreting and maturing the recombinant protein in mammary tissue;
- b. collecting the milk; and
- c. isolating the exogenous recombinant protein from the milk.

33. The process according to claim 32, wherein said expression system also includes a 3' flanking region coding for  $\alpha$ -lactalbumin downstream of the DNA sequence.

34. The process according to claim 32, wherein said expression system also includes a 5' flanking region

-36-

coding for the signal peptide upstream of the DNA sequence.

35. A selection characteristic for identifying superior milk producing mammals comprising a DNA sequence encoding bovine  $\alpha$ -lactalbumin and having the DNA sequence (SEQ ID NO:3) listed in Fig. 5 in the control region of bovine  $\alpha$ -lactalbumin.

36. A selection characteristic for identifying superior milk producing mammals comprising inherited genetic material which is a mammary specific DNA sequence encoding bovine  $\alpha$ -lactalbumin and promoting quantitative differences in gene expression among mammals, wherein the DNA sequence is characterized by variations in the gene structure in the control region of bovine  $\alpha$ -lactalbumin.

37. The selection characteristic of claim 36 wherein one of the variations is in the -13 position of the DNA sequence.

38. The selection characteristic of claim 37 wherein the -13 position is occupied by adenosine.

39. The selection characteristic of claim 36 comprising the following DNA sequence (SEQ ID NO:19) in the control region of bovine  $\alpha$ -lactalbumin:  
TATAAGAAATCAGGCTTTAGAGACTGATGTAGAGAGAATGAGCCCTGGCATAACCAGA  
AGCTAACAGCTA.

40. The selection characteristic of claim 36 comprising the following DNA sequence (SEQ ID NO:20) in the control region of bovine  $\alpha$ -lactalbumin:  
AGGAAGCTCAATGTTTCTTTGTTGGTTTTACTGGCCTCTCTTGTC.

41. The selection characteristic of claim 36 comprising the DNA sequence (SEQ ID NO:3) listed in Fig. 5 in the control region of bovine  $\alpha$ -lactalbumin.

42. A method of predicting superior milk and milk protein production in mammals comprising comparing selected positions on the DNA sequence of the inherited control region for  $\alpha$ -lactalbumin in a subject mammal with analogous positions on the DNA sequence of the

-37-

control region for  $\alpha$ -lactalbumin from mammals known for superior milk and milk protein production.

43. The method of claim 42 wherein one of the selected positions is the -13 position on the control region of the DNA sequence.

44. The method of claim 43 wherein the -13 position is occupied by the base adenine.

45. The method of claim 42 wherein the selected DNA sequence comprises a steroid response element.

## AMENDED CLAIMS

[received by the International Bureau on 14 December 1992 (14.12.92);  
original claims 2,5,9-11,19,24-31,33,35-45 cancelled;  
original claims 1,3,6,8,12,13,15,17,18,20-23,32 amended;  
other claims unchanged (4 pages)]

1. An isolated DNA sequence which promotes mammary specific expression of mRNA in mammary cells of lactating animals comprising a variant of bovine  $\alpha$ -lactalbumin 5' flanking region regulatory sequence wherein the variant is in the -13 position from the start of the signal peptide coding region.

3. The DNA sequence of claim 1 wherein the -13 position is occupied by adenine.

4. The DNA sequence of claim 1 comprising DNA sequence (SEQ ID NO:19) in the control region of bovine  $\alpha$ -lactalbumin: TATAAGAAATCAGGCTTTAGAGACTGATGAGAGAAATGAGCCCTGGCATACCAAGCTA ACAGCTA.

6. The DNA sequence of claim 1 comprising the DNA sequence listed in Fig. 5 (SEQ ID NO:3) as the 5' flanking region of bovine  $\alpha$ -lactalbumin.

7. An expression vector comprising the DNA sequence of claim 1.

8. An expression system comprising a mammary specific  $\alpha$ -lactalbumin control region construct operatively linked to a DNA sequence coding for a signal peptide, which when genetically incorporated into a non-human mammal permits the female species of that mammal to produce the desired recombinant protein in its milk.



12. The expression system of claim 8 wherein the construct comprises a 5'  $\alpha$ -lactalbumin flanking region attached to a bovine  $\beta$ -casein DNA sequence.

13. The expression system of claim 12 wherein the construct contains the polyadenylation site of  $\beta$ -casein 5' flanking region.

14. The expression system of claim 12 wherein the construct includes the proximal promoter from a first milk protein and the distal control region of a second milk protein.

15. The expression system of claim 14 wherein the first and second milk proteins are selected from the group consisting of  $\alpha$ -lactalbumin,  $\beta$ -lactoglobulin,  $\beta$ -casein,  $\alpha$ <sub>1</sub>-casein,  $\alpha$ <sub>2</sub>-casein and K-casein.

16. The expressions system of claim 12 wherein the construct includes the proximal promoter of  $\beta$ -casein and the distal control region of  $\alpha$ -lactalbumin.

17. A transgenic non-human mammal containing an expression system comprising the DNA sequence listed in Fig. 5 (SEQ ID NO:3) as the  $\alpha$ -lactalbumin control region operatively linked to an exogenous DNA sequence coding for a desired protein to be expressed in milk through a DNA sequence coding for a signal peptide effective in secreting and maturing the protein in mammary tissue.

18. The transgenic non-human mammal of claim 17 wherein the  $\alpha$ -lactalbumin control region includes a mammary specific DNA sequence encoding bovine  $\alpha$ -lactalbumin and having the following nucleotide sequence (SEQ ID NO:19) in the control region of bovine  $\alpha$ -lactalbumin:

TATAAGAAATCAGGCTTTAGAGACTGATGTAGAGAGAATGAGCCCTGGCATAACAGAAGCTA  
ACAGCTA.

20. Products produced by the genetically engineered non-human mammal of claim 17.
21. Semen produced by the transgenic non-human mammal of claim 17.
22. Milk produced by the transgenic non-human mammal of claim 17.
23. A transgenic non-human mammal of claim 17.

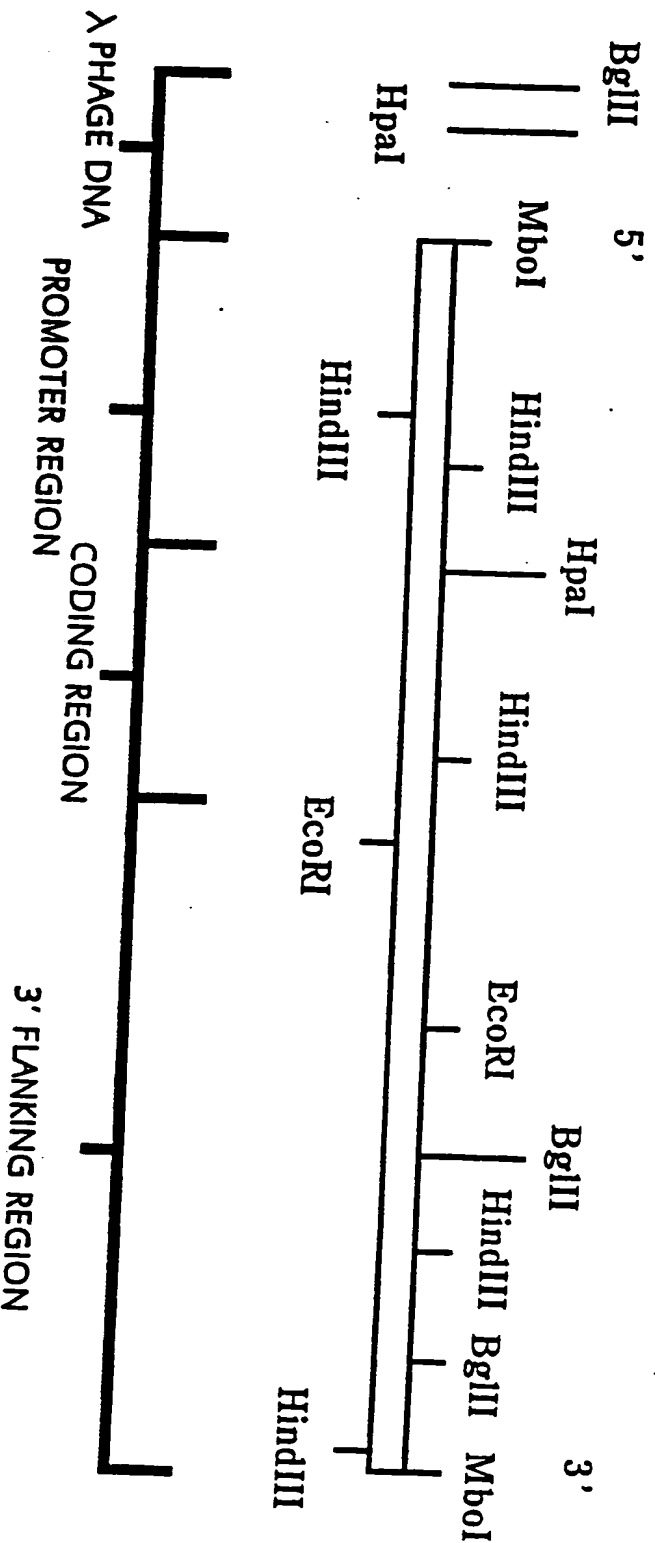
32. A process for the production and secretion into non-human mammal's milk of an exogenous recombinant protein comprising the steps of:

- a. producing milk in a genetically engineered mammal containing an expression system comprising a mammary specific  $\alpha$ -lactalbumin control region operatively linked to a DNA sequence coding for a signal peptide, which is effective in secreting and maturing the recombinant protein in mammary tissue;
- b. collecting the milk; and
- c. isolating the exogenous recombinant protein from the milk.

34. The process according to claim 32, wherein said expression system also includes a 5' flanking region coding for the signal peptide upstream of the DNA sequence.

ALPHA-LACTALBUMIN GENOMIC CLONE: FIG. 1

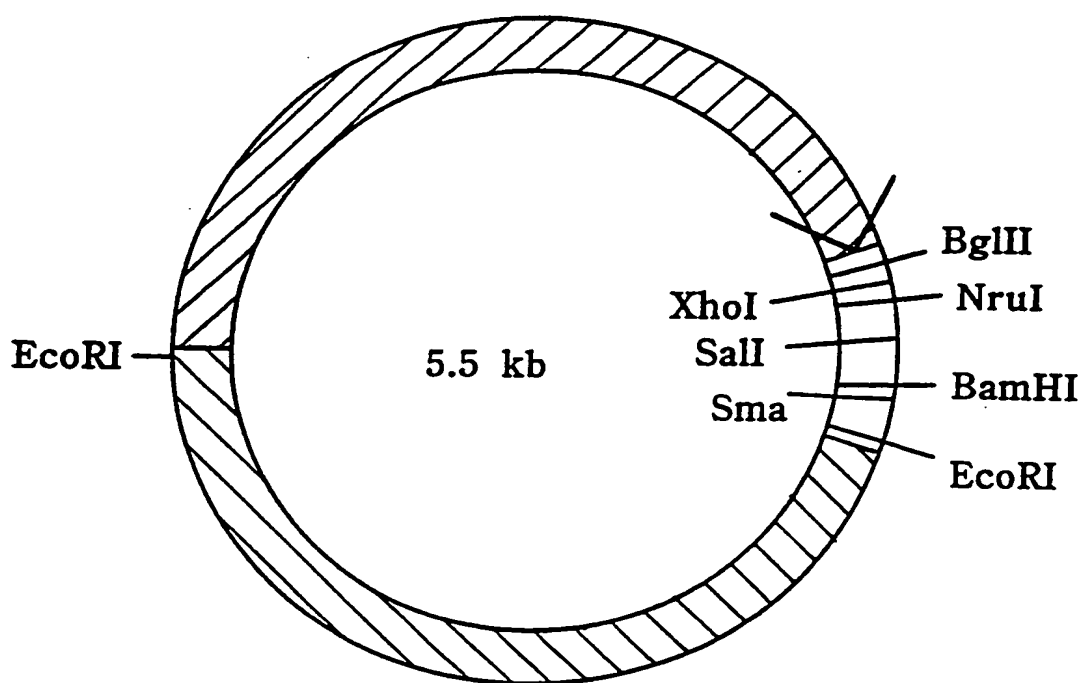
2.0 kb 5' FLANKING REGION (PROMOTER REGION)  
 1.7 kb CODING REGION (4 EXONS)  
 8.8 kb 3' FLANKING REGION



2/14

FIG. 2

A-lac Pro/pIC 20R



  $\alpha$ -LACTALBUMIN PROMOTER REGION

 pIC 20R PLASMID SEQUENCE

 MULTIPLE CLONING SITE

# Alpha-lactalbumin Promoter Region Cloned in EcoRV Site of Pic 20R:

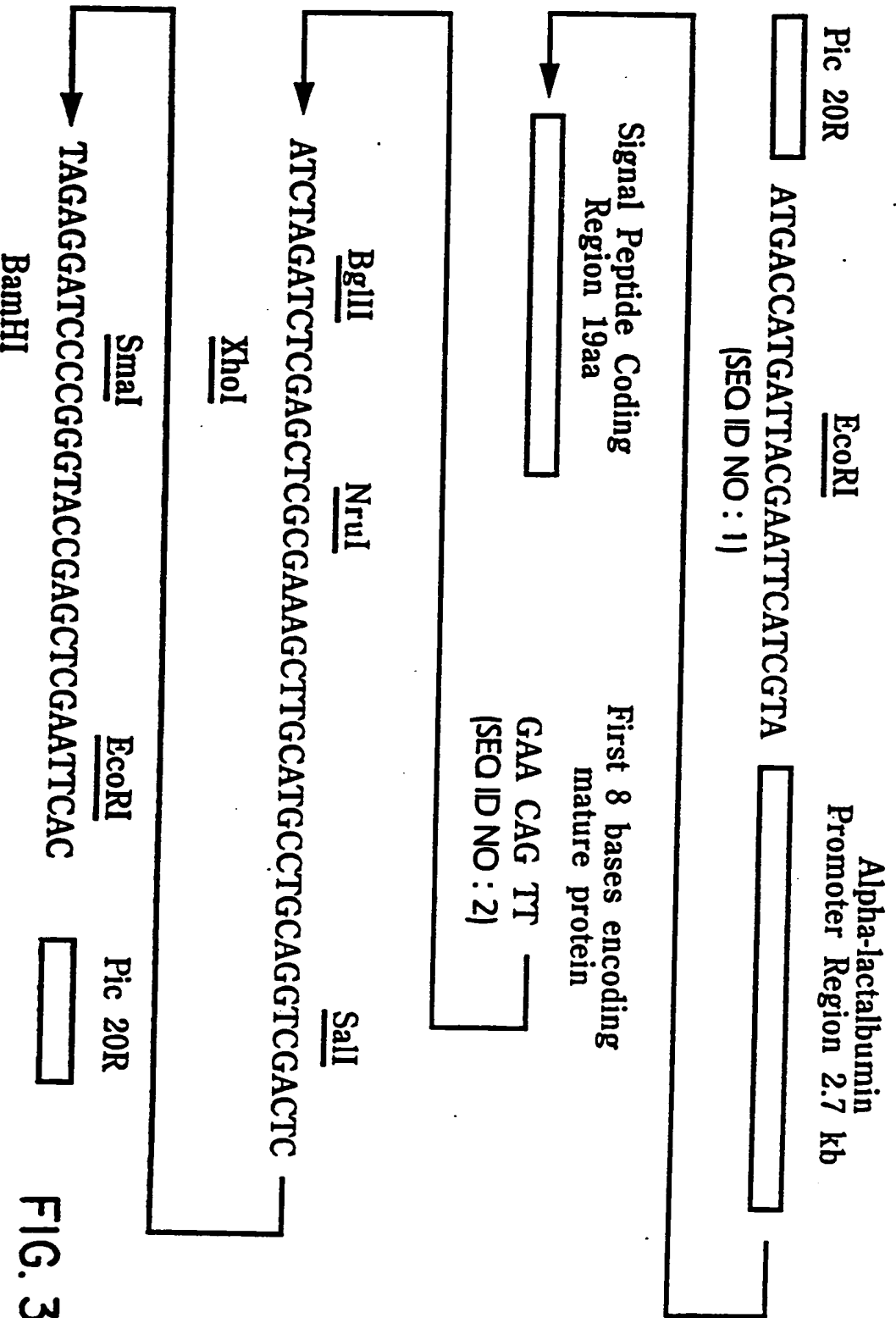
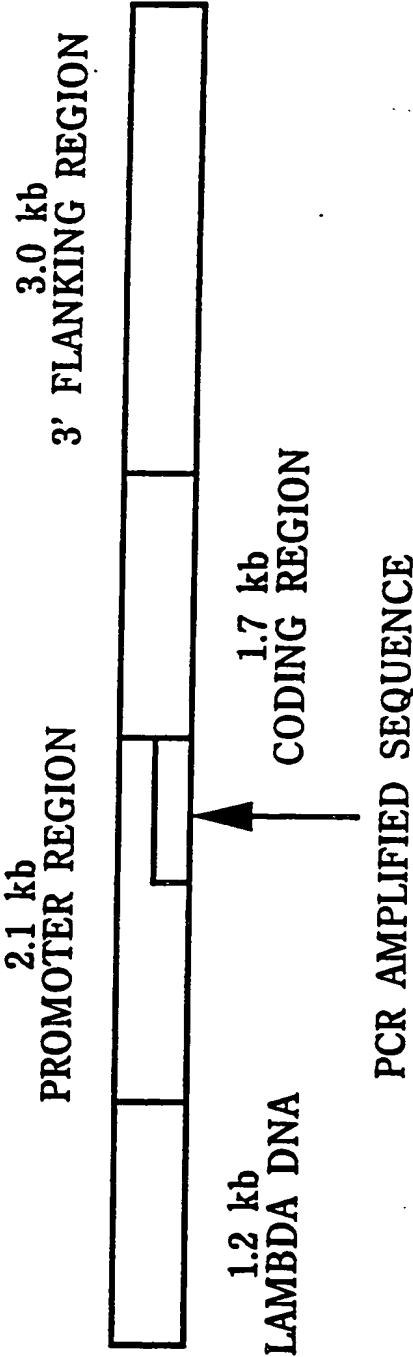


FIG. 3

SUBSTITUTE SHEET

FIG. 4



SEQ ID NO:3

FIG. 5A

5/14

1 GATCAGTCT GGGTGGTCAI TGAAAGGACT GATGCTGAG TTGAGCTCC AATACTTGG  
 CCACCTGATG CGAAGAACTG ACTCATGTGA TAGACCCCTG  
 101 AATCTGGGA AGATTGAAG CAGGAGGAGA AGGGATGACA GAGGATGGA GAGTTGATG  
 GAATCACCMA CTCGATGGAC ATGAGTTTGA GCMAGCTTCC  
 201 AGGAGTTGT AATGGGCAGG GAGCCCTGGC GTGCTGCAGT CCAITGGGGT GCMAAGATT  
 GGACACTACT GAGTGACTGA ACTGAACTGA TAGTGAATC  
 301 CATGGTACAG AATAAGGAT AAAAAGAGG AGAGTTTGC CCTGATCTG AGAGTTGTA  
 GGATAIAMA GTTAGATA CTTTAGTTT GGAGTCTTA  
 401 AATTATTAC TTAGGATGG TACCCACTGC AATAAGAA ATCAGGCTTT AGAGACTGAT  
 GTAGAGAGAA TGAGCCCTGG CATACCAGAA GCTAACAGCT  
 501 AITGGTAIA GCTGTAIA CCAATAIAIA ACCAAIAIAT TGGTAIAIA GCATGAAGCT  
 TGATGCCAGC AATTGAAGG ACCAATTAG AACTAGTAC  
 601 CTAACTCTA CAITGCCAG GACACTGAATC TTAAAGCTCA GGTTCAGAAI CTGTTTAT  
 AGGCTCTAGG TGAIAITGT GGGGCTTCCC TGGTGGCTCA  
 701 GATGGTAAG TGTCTGCTG CAITGTGGT GATCTGGGT CGATCCCTGG CTGGGAGA  
 TCCCTGGAG AAGGAATGG CAACCCACTC TAGTACTCT  
 801 ACCTGGAAA TTCATGAC AGAGGAGCTT TGTAGCTAC AGTCCATGGG ATTGCMAAGA  
 GTTGACACA ACTGAGCAC TAGCACAGC ACAGTACAGT  
 901 AATACCTGT GAGGTGAAGT GAAGTGAGG TTCAITGAC GGTCTCTGCT ATTGCAGAA  
 GATCTTAC CAITGAAGC ACCAGGAGG CCAAGATA

SEQUENCE SHEET



6/14

## FIG. 5B

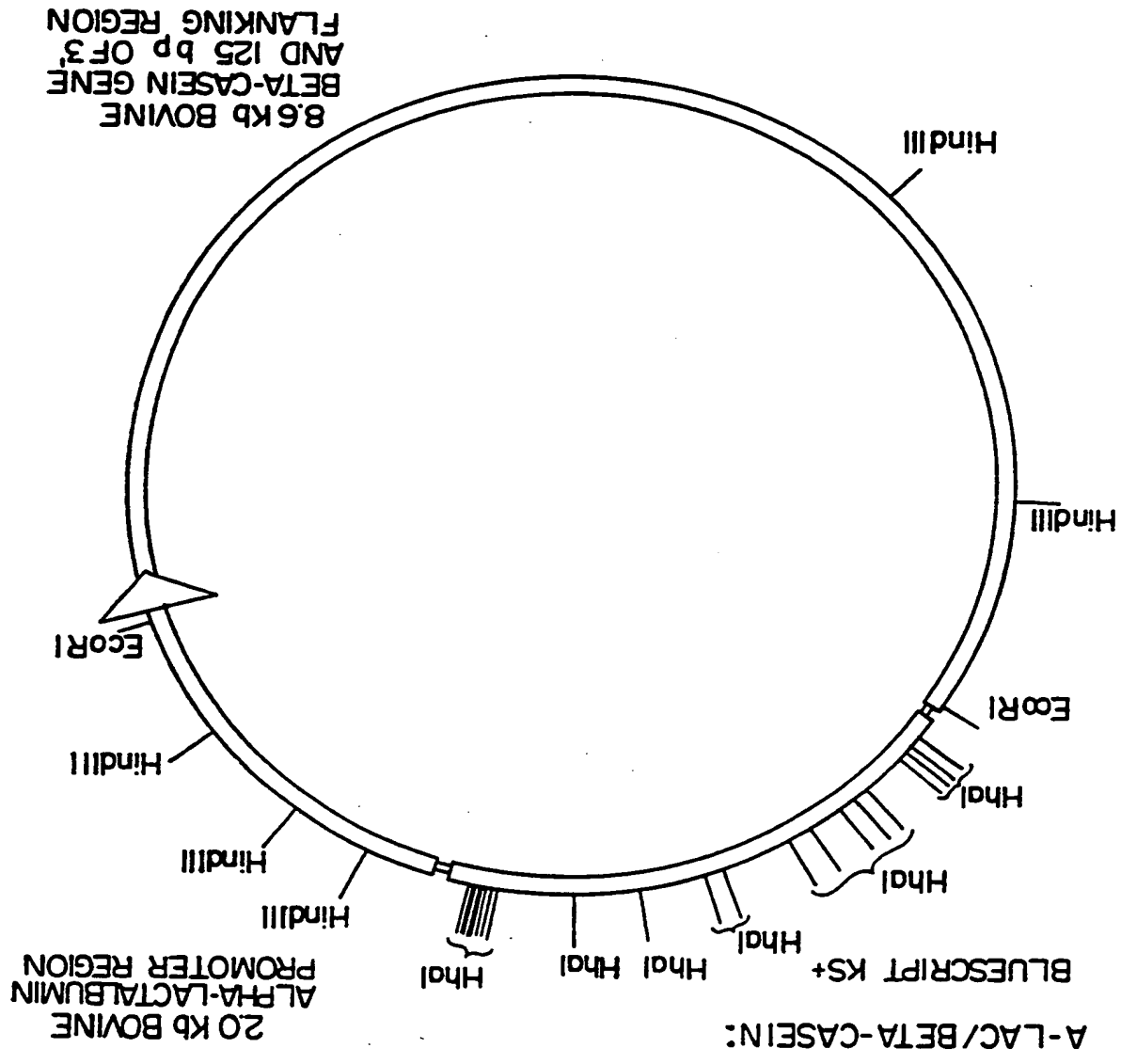
SEQ ID NO : 3

```

1001 CTGGAGTGGG TAGCCTATTC CTCTCCAGG GGATCTTCCC ATCCCAGGAA TTGAACTGGA
    GTCTCTGCA TTCAGGTGG ATCTTCACC AGCTGAAC TA
1101 CCAGGTGGAT ACTACTCCAA TATTAAAGTG CTTAAAGTCC CCTTCCCAA
    AAGGTTGGG TCACCTTTT TTAACCTTCT GTGGCCTACT
1201 CTGAGGCTGT CTACAGCTT ATATATTAT GAACACATTT GTTAGTTTA
    GATTACAAT GTGGTATCTG GCTATTAGT GGTATTGGTG
1301 GTTGGGGATG GGGAGGCTGA TAGCATCTCA GAGGGCAGCT AGATACTGTC ATACACACTT
    TTCAAGTTCT CCATTTTGT GAAATAGAA GTCTCTGGAT
1401 CTAAGTTATA TGTGATTCTC AGTCTCIGTG GTCATATTCT TGACCACTCA
    ACAAGGAACC AAGATATCAA GGGACACTTG TTTTGTTTCA
1501 TGCCTGGGTT GAGTGGGCCA TGACATAIGA TGATGTACAG TACTCTGTAT
    GTCTCTAAGA GGAAGGAGGA GTTGGCCGTG GACCCTTTGT
1601 GCATTTTCTG ATTGCTTCAC TTGTATTACC CCTGAGGCCC TGAATAGGT
    TGGGCACATC TTGCTTCCTA GAACCAACAC TACCAGAAC
1701 AACATAATA AAGCCAATG GGAAACAGGA TCATGTTTGT GGGCAGGTAA
    CAATACCTAG TATGGACTAG AGATTCTGGG GAGGAAGGA
1801 AAGTGGGGT GAATTACTG AAGGAAGTCT ATGTTTCTT ACTGGCCTCT
    CTTGTCATCC TCTTCTGGA TGTAAAGGCTT GATGCCAGGG
1901 CCCCTAAGGC TTTTCCACA ATAAAGGA GGTGAGCAGT CATTCAGAA
    TCTTGAGGGG TACCAAAAT GATGTCCTT GTCTCTCTGC
2001 TCCTGGTAGG CATCTATTC CATGCCACCC AGGCTGAACA GTTA

```

FIG. 6



8/14

### Putative Steroid Response Element

Bovine (Our Sequence) [-546]-[-533]: CATATTCTATT.CTA (SEQ ID NO:4)  
 |||||\*|||  
 Human (Hall) [-587]-[-574]: CATATTCTATTCCCTA (SEQ ID NO:5)  
 |||||\*|||  
 Bovine (Vilotte) [-547]-[-533]: CATATTCTATTCTA (SEQ ID NO:6)  
 |||||\*|||

### RNA Polymerase Binding Region

Bovine (Our Sequence) [-18]-[-1]: TCTTGAGGGGTAACCAAA (SEQ ID NO:7)  
 |||||\*|||\*|||  
 Human (Hall) [-17]-[-1]: TCTTG.GGGTAGCCAAA (SEQ ID NO:8)  
 |||||\*|||\*|||  
 Bovine (Vilotte) [-18]-[-1]: TCTTGAGGGGTCACCAAA (SEQ ID NO:9)  
 |||||\*|||\*|||

FIG. 7

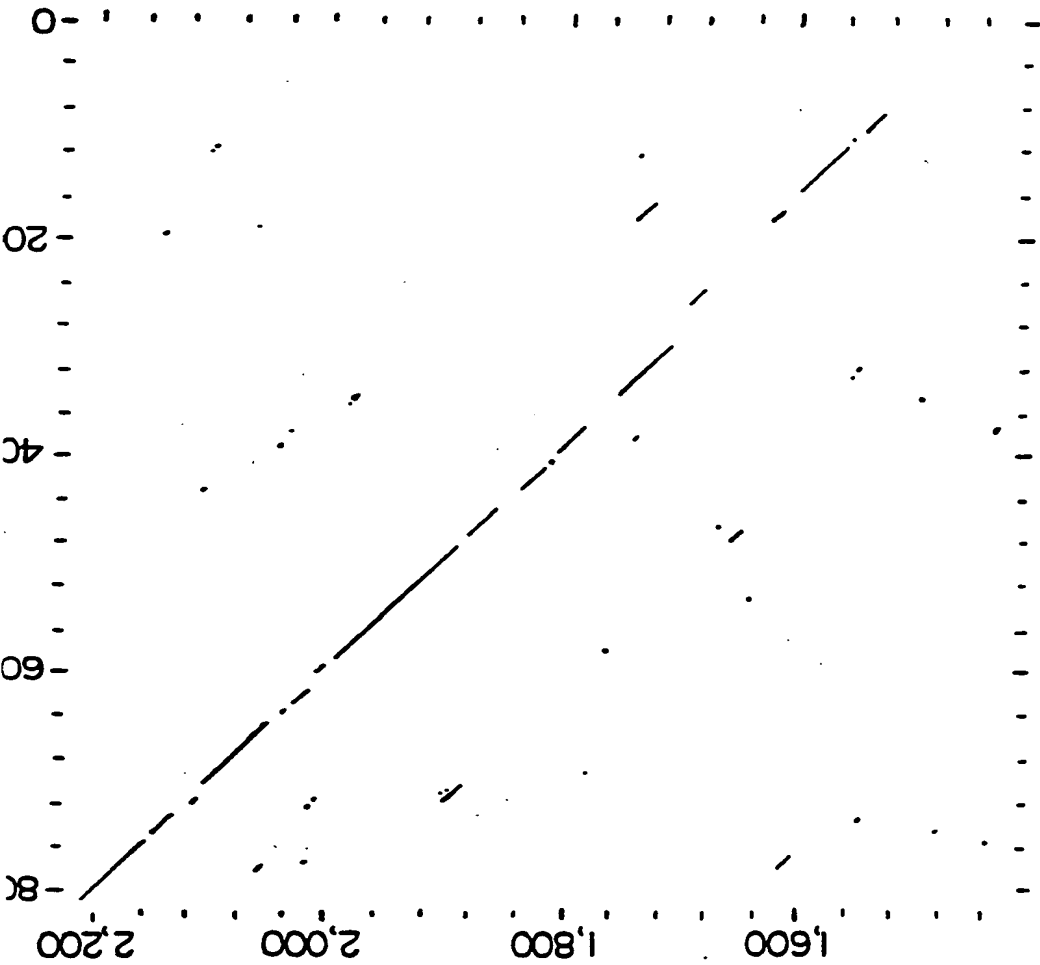


9/14

FIG. 8

BOVINE ALPHA-LACTALBUMIN 5' FLANKING REGION

HUMAN ALPHA-LACTALBUMIN 5' FLANKING REGION



10/14

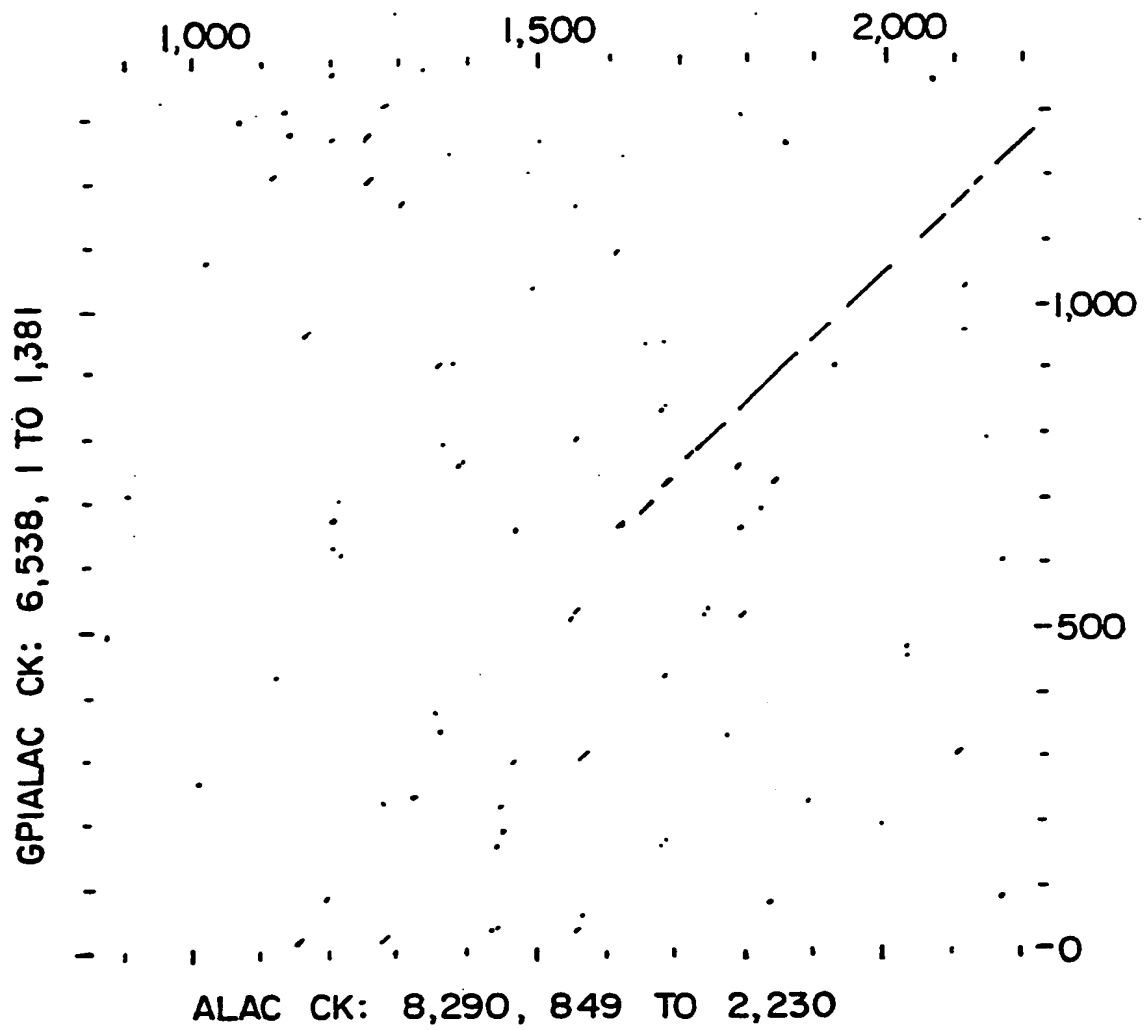
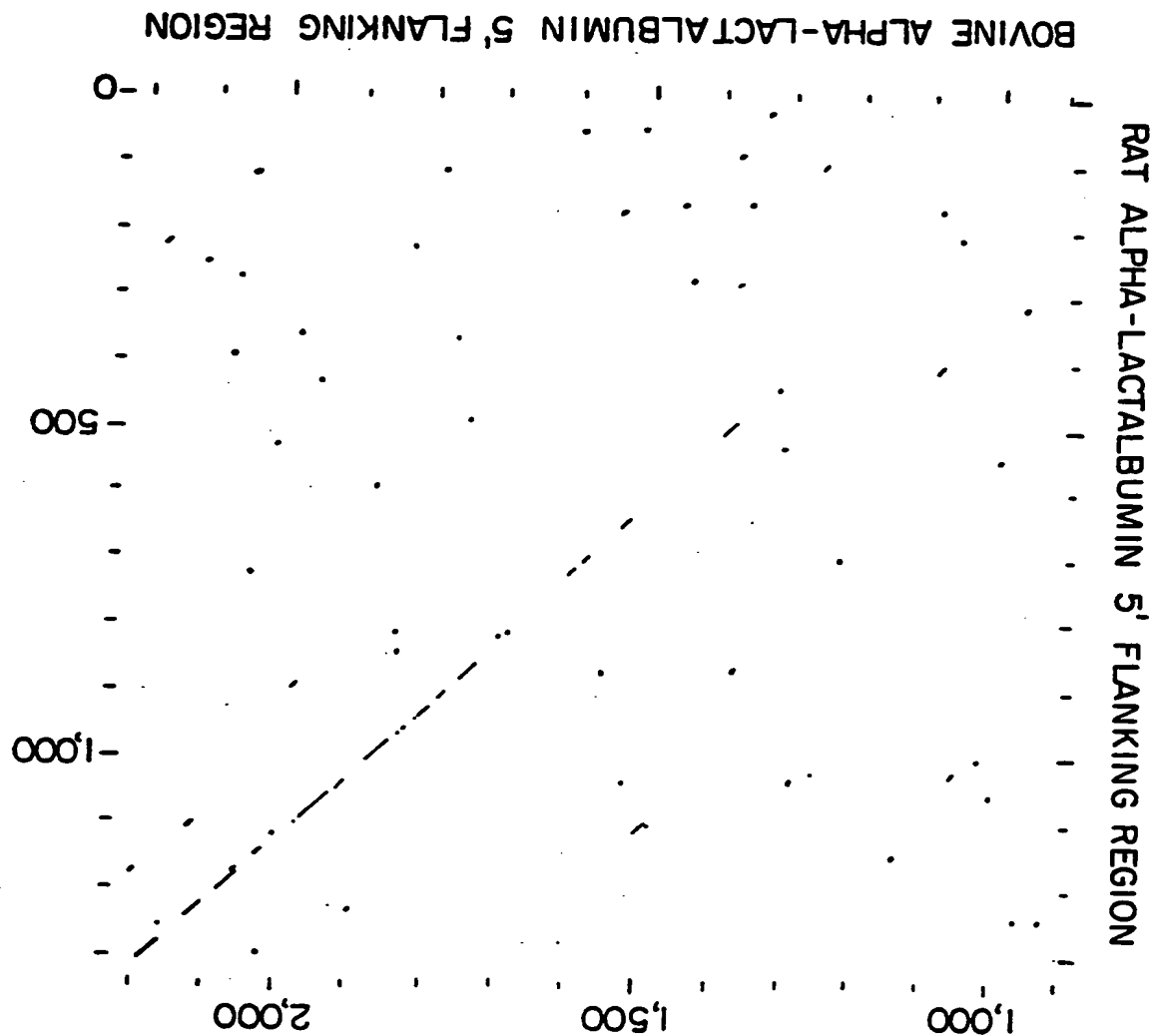


FIG. 9

FIG. 10



11/14

12/14

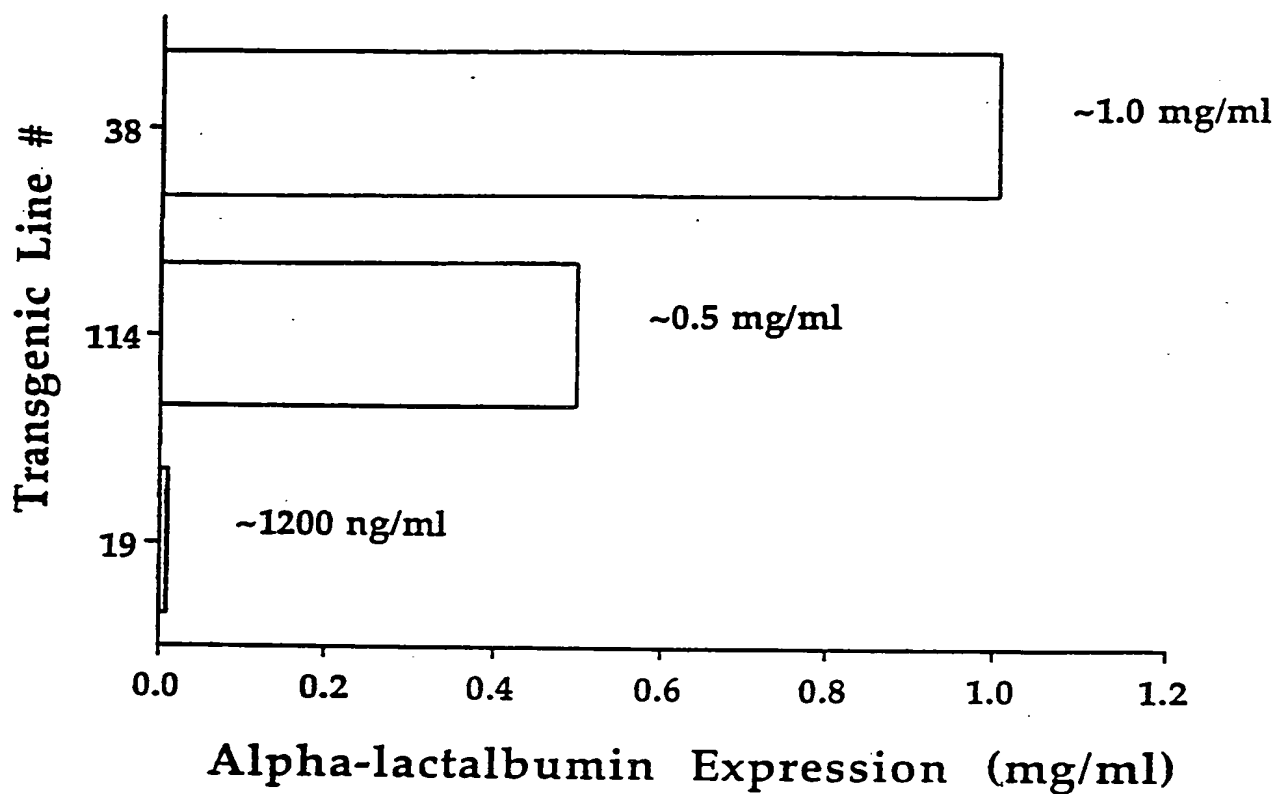
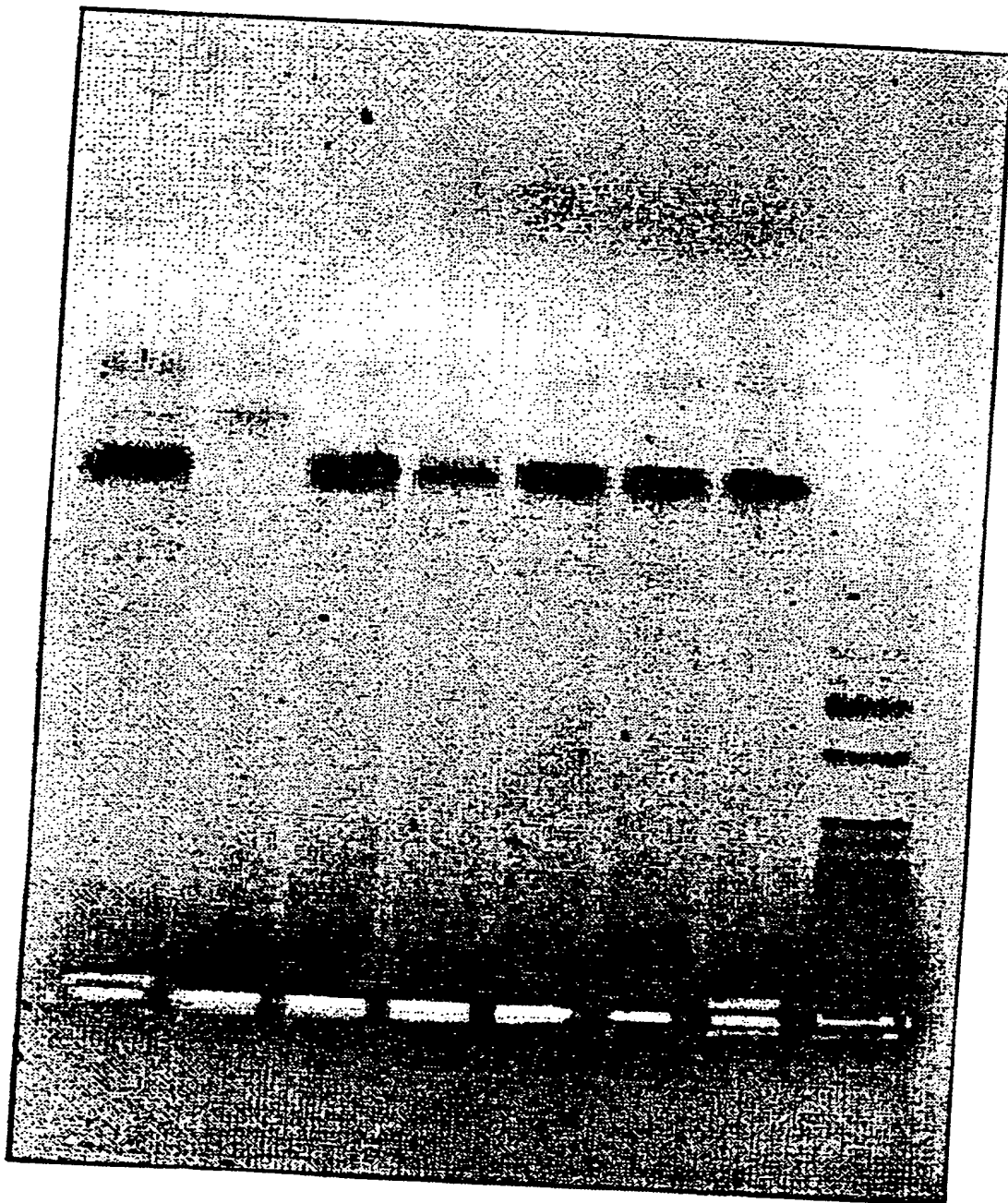


FIG. II

FIG. 12





14/14

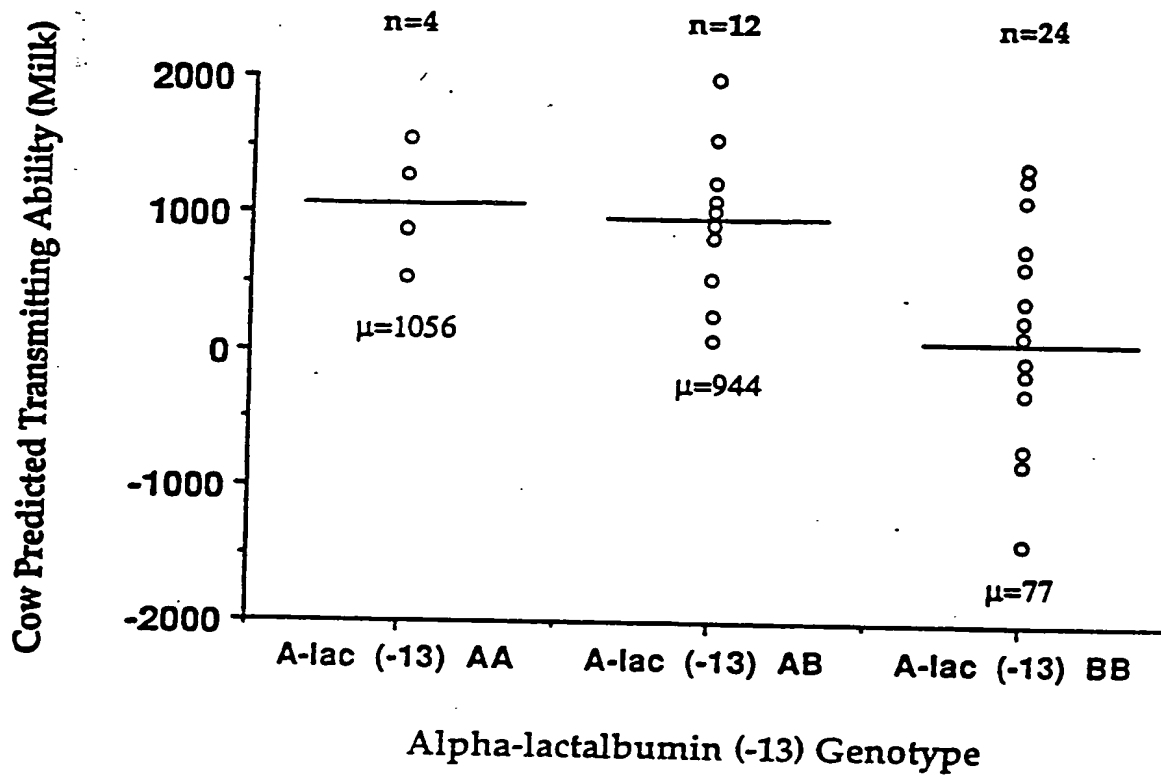


FIG.13

**International application No.**

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(5)  
US CL: 800/2; 435/172.3, 69.1; 935/6, 63  
Please See Extra Sheet.

**B. FIELDS SEARCHED**

Minimum document length searched (classification system followed by classification symbols)

U.S. : 800/2; 435/172.3, 69.1; 935/6, 63

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

## APS, CAS Online

Execution date based on the information and, where practicable, search results:

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to c
-----------	--	---------------

1-2, 5, 7-11 24, 26, 28-30 and entire document.	European Journal of Biochemistry, Volume 186, issued 1989, J-L. Vilotte et al., "Efficient Tissue-Specific Expression of Bovine $\alpha$ -Lactalbumin in Transgenic Mice", pages 43-48, see entire document.
3, 4, 6 and 35-45	Biochemical Journal, Volume 242, issued 1987, D. C. Hall, "Organization and Sequence of the Human $\alpha$ -Lactalbumin Gene", pages 735-742, see entire document.
12-16	Nucleic Acids Research, Volume 16, No. 3, issued 1988, K-F. Lee et al., "Tissue specific Expression of the Rat B-Casein Gene in Transgenic Mice", pages 1027-1041, especially 1033-1037.

Further documents are listed in the continuation of Box C. ☐ See patent family annex. ☐

- | Special categories of cited documents:   | T | X | Y | Z |
|--|---|---|---|---|
| documents defining the general state of the art which is not considered to be part of particular reference   |   |   |   |   |
| earlier documents published on or after the international filing date  |   |   |   |   |
| documents which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) |   |   |   |   |
| documents referring to an oral disclosure, use, exhibition or other means  |   |   |   |   |
| documents published prior to the international filing date but later than the priority date claimed  |   |   |   |   |

State of the actual completion of the international search

19 September 1992

same and mailing address of the ISA/  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231  
estmille No. NOT APPLICABLE

**Authorized officer**

Telephone No. (703) 308-1126

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US92/06549

A. CLASSIFICATION OF SUBJECT MATTER:

IPC (5):

C12N 15/00; C12P 21/06.

**THIS PAGE BLANK (USPTO)**